REPORT RESUMES

ED 002 965 24

INVESTIGATIONS OF THE CHARACTERISTICS OF PROGRAMED LEARNING SEQUENCES.

BY- GLASER, ROBERT AND OTHERS
PITTSBURGH UNIV., PA.

REPORT NUMBER CRP-691 PUB DATE OCT 61
EDRS PRICE MF-$0.27 HC-$7.72 193P.

DESCRIPTORS- EDUCATIONAL RESEARCH, BEHAVIOR, STIMULUS BEHAVIOR, *PROGRAMED INSTRUCTION, PROGRAMING, *LEARNING THEORIES, REINFORCEMENT, *AUTOINSTRUCTIONAL AIDS, *AUTOINSTRUCTIONAL METHODS, *AUTOINSTRUCTIONAL PROGRAMS, PROGRAMED TEXTS, PHONOTAPE RECORDINGS, PITTSBURGH, PENNSYLVANIA

THIS STUDY SOUGHT TO IDENTIFY THOSE PROGRAM CHARACTERISTICS WHICH PRODUCE MAXIMUM LEARNING WITH SELF-TUTORING DEVICES. AIMS OF THE 2-YEAR PROJECT WERE (1) TO DEVELOP PROCEDURES FOR ANALYZING SUBJECT MATTER FOR CONVERTING IT INTO PROGRAMED LEARNING SEQUENCES, AND (2) TO CARRY OUT EXPERIMENTAL COMPARISONS BETWEEN SELECTED VARIATIONS OF PROGRAMED SEQUENCES. THE STUDY WAS COMPOSED OF 9 SEPARATE EXPERIMENTAL PROJECTS CONDUCTED OVER A 2-YEAR PERIOD. THEY ARE (1) DEVELOPMENT OF THE RULEG SYSTEM FOR THE CONSTRUCTION OF PROGRAMED VERBAL LEARNING SEQUENCES, (2) AN INVESTIGATION OF TEACHING MACHINE VARIABLES USING LEARNING PROGRAMS IN SYMBOLIC LOGIC, (3) A STUDY OF THE EFFECTS OF PARTIAL REINFORCEMENT ON BEHAVIOR IN A PROGRAMED LEARNING SITUATION, (4) A COMPARISON OF THREE METHODS OF PROGRAMING SUBJECT MATERIAL FOR AUTOINSTRUCTION, (5) THE APPLICATION OF PROGRAMED LEARNING TECHNIQUES TO SELECTED MATERIAL IN A PROFESSIONAL CURRICULUM (A STUDY OF RETENTION EFFECTS), (6) AN INVESTIGATION OF VISUAL VERSUS AUDITORY PROGRAMING IN TEACHING FOREIGN LANGUAGE PRONUNCIATION, (7) DEVELOPMENT OF A VOCABULARY PROGRAM USING LANGUAGE REDUNDANCY, (8) AN EXPLORATORY EVALUATION OF A DISSEMINATIVE TRANSFER-LEARNING PROGRAM USING LITERAL PROMPTS, AND (9) DEVELOPMENT OF SYMBOLS FOR USE IN THE EDITING OF PROGRAMED LEARNING SEQUENCES. THE IMPLICATIONS OF THE PROJECT POINTED OUT FIVE RESEARCH PROBLEM AREAS IN PROGRAMED LEARNING. (HB)

691

ED002965

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

INVESTIGATIONS OF THE CHARACTERISTICS OF PROGRAMMED LEARNING SEQUENCES

Robert Glaser, Project Director
Julian I. Taber, Associate Project Director

Programmed Learning Laboratory
Department of Psychology
University of Pittsburgh

INVESTIGATIONS OF THE CHARACTERISTICS OF PROGRAMMED LEARNING SEQUENCES

Cooperative Research Project No. [redacted]
691

Robert Glaser, Project Director

Julian I. Taber, Associate Project Director

Contributors

Attila P. Csanyi
James L. Evans
Carole Finelli
Robert Glaser
Robert B. Hessert
Lloyd E. Homme
Halmuth H. Schaefer
Eugenia S. Scharf
Julian I. Taber

Programmed Learning Laboratory
Department of Psychology
University of Pittsburgh
Pittsburgh 13, Pennsylvania

October 1961

The research reported herein was supported through the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

Copyright 1961 by the University of Pittsburgh

3414

Acknowledgements

It is impossible to acknowledge all of those who provided us with stimulation and feedback during the course of the project. People particularly concerned with the separate studies described in this report are acknowledged at the beginning of each chapter. We wish to express special appreciation to the Cooperative Research Branch of the U.S. Office of Education for their wise administration and monitoring of the efforts of researchers. Dr. Robert A. Patton, Head of the Department of Psychology of the University of Pittsburgh, provided always helpful administrative guidance. Dr. J. Steele Gow, Jr., Director of the Coordinated Education Center of the University of Pittsburgh, constantly encouraged our efforts. Margaret G. Fullick, technical assistant to the Director of the Programmed Learning Laboratory, provided many valuable services throughout the project. The excellent secretarial skills and services of Theresa Boney, Ruth Ann Rishel, and Judith Wedmore are sincerely appreciated.

October 1961

Robert Glaser
Julian I. Taber

TABLE OF CONTENTS

		Page
Chapter 1.	Introduction, Summary, and Implications	1
Chapter 2.	The Ruleg System for the Construction of Programmed Verbal Learning Sequences	17
Chapter 3.	An Investigation of "Teaching Machine" Variables Using Learning Programs in Symbolic Logic	31
Chapter 4.	A Study of the Effects of Partial Reinforcement on Behavior in a Programmed Learning Situation	83
Chapter 5.	A Comparison of Three Methods of Programming Subject Material for Auto-Instruction	97
Chapter 6.	The Application of Programmed Learning Techniques to Selected Material in a Professional Curriculum: A Study of Retention Effects	115
Chapter 7.	An Investigation of Visual Versus Auditory Programming in Teaching Foreign Language Pronunciation	135
Chapter 8.	A Vocabulary Program Using "Language Redundancy"	151
Chapter 9.	An Exploratory Evaluation of a Discriminative Transfer Learning Program Using Literal Prompts	163
Chapter 10.	Symbols for Use in the Editing of Programmed Learning Sequences.	177

Chapter 1

Introduction, Summary and Implications[1]

The Aims of the Project

When work was begun on this project in the fall of 1959, the field of teaching machines and programmed and spelling learning was experiencing its initial rapid growth. At that time, there existed virtually no course-length programmed learning sequences. Many psychologists and engineers diverted their energies to the construction of new devices for the display of programs, and workers in the field, for the most part, were only beginning to discover that by far the most important aspect of the new instructional technology was the learning program, certainly not the teaching machine. The psychological literature contained no extensive and specific recommendations on how programs should be constructed. What recommendations existed were in the form of relatively intuitive generalizations from experiments in the learning laboratory. Conflicts existed between methods which reflected, primarily, the lack of research in the applied instructional setting. Skinner was recommending short frames, constructed answers, and immediate confirmation. Pressey had recommended that the primary learning task should take place in the usual lecture-type and textbook situation and that learning could be enhanced by the use of multiple-choice answer questions which received immediate confirmation. Crowder was advocating the use of long textual passages followed by a single multiple choice question, items which appeared in "scrambled textbooks" or on film in automated devices. Skinner's notions were of greatest appeal to psychologists because of the success of his concepts in the learning laboratory. Practical programmers, often non-psychologists, tended to concentrate on the superficial differences between these methods. The experimentally minded went to work realizing that the proposed techniques were provocative, but crude, beginnings; research and development were required and future programmed learning sequences might look much different from the first attempts.

---

[1] This chapter was prepared by Robert Glaser, Project Director, and J. I. Taber, Associate Project Director.

In brief, two years ago the field of auto-instructional techniques was characterized by a lack of constructed and tested programs, an overabundance of teaching machines, confusing methodological claims, and intensive growth. While during the past two years the theoretical and methodological arguments, if anything, have increased in intensity and scope, the number of available and tested programs has increased, and the relative importance of teaching machines as opposed to learning programs has become clear for most workers in the field. Whereas, two years ago much effort was being devoted to the automation of learning, now most research and development efforts in the field are devoted to the construction of an environment in which learning is highly probable, i.e., to the construction of learning programs which can be presented either in programmed textbook format or in automated devices.

The instigators of the present project,[2] in the fall of 1959, selected certain aspects of the field towards which to devote their energies. From its inception, this project has been concerned with two major objectives in the area of programmed learning, objectives which lie within the region of overlap between basic research in the behavioral sciences and educational practice. The aims of the project were: (a) to attempt to develop procedures for analyzing subject matter for converting it into programmed learning sequences, and (b) to carry out experimental comparisons between selected variations of programmed sequences. In view of the rapidity of change taking place in the field of programmed instruction, a philosophy of research was adapted which permitted a variety of research endeavors to progress simultaneously rather than one which committed project personnel to a single, experimental design. The final project report consists of 10 chapters including the introduction. These 10 chapters summarize the work of project personnel over a two-year period. Each chapter has been cast in the form of a technical report of research, that is, each chapter is in itself a final report of a particular experimental project. The reports themselves deal with a wide range of problems in programmed learning. This work has been devoted to such aspects of auto-instructional techniques as the properties of the learning program itself, the nature of the student, the nature of the subject matter, methods of programming the subject matter, and methods of training non-psychologists in the construction and use of programs.

---

[2] From September 1959-August 1960, Robert Glaser and Lloyd E. Homme were co-directors; from September 1960-August 1961, Dr. Glaser was Project Director and Julian I. Taber, Associate Project Director.

In light of the absence of a technology for converting subject matter into programmed sequences, project personnel first turned their attention to "the analysis of knowledge." This was in line with the first objective of the project. The development of a good teaching program requires the elaboration of those specific behaviors which reflect, or are said to reflect, "knowledge of subject matter." Techniques for extracting and elaborating these specific behaviors, behaviors which must be built into a program, were an unspecified aspect of programming procedure. Chapter 2 of this report presents a generalized scheme for the analysis of subject matter for programming purposes. Subsequent sections of the final report describe work attempts on the second objective of the project, i.e., investigations of variations in the properties of programmed learning sequences which might affect learning. Chapters 3, 4, and 5 represent such investigations of programmed learning sequences. These chapters also represent the experimental application of the Ruleg Programming System described in Chapter 2, and they use a specialized program related to symbolic logic, which was originally based upon Ruleg programming. Chapters 6 and 7, while they continue the parametric investigations of programming variables, also represent attempts at the implementation of programmed learning in actual instructional settings. Chapters 6 and 7 deal with actual subject matters of interest to teachers and represent studies carried out "in the field."

The work on the two major aims of the project led to two other identifiable interests on the part of project personnel. One of these interests is represented in Chapters 8 and 9. These chapters are concerned with developing new programming techniques. The types of programs suggested in these two chapters, as well as that suggested in Chapter 7 on teaching Spanish pronunciation, are programs which are dissimilar to the types of programs generally being considered in current programming work. In the course of the project the examination of basic behavioral principles led to innovations in programming techniques, innovations which have not yet been widely tested and evaluated. Thus, some of the chapters in this final report reflect rather informal developmental attempts. The other subsidiary interests revolved around attempts to train non-psychologists in the technology of programming. Chapter 10, particularly, represents this interest. The work presented in the remaining nine chapters, then, represents both research and development.

The following section of this introduction consists of a series of summaries of each of the individual chapters. These summaries will be necessarily brief and, although results of each of the studies may seem in summary to be rather clear-cut, the reader is cautioned that safe conclusions are drawn only from a detailed consideration of the complete reports. The third and final section of this introductory chapter presents an overview of the major implications of the total work of the project.

Summaries of the Studies Carried Out.

Chapter 2: The Ruleg System for the Construction of Programmed Verbal Learning Sequences. This chapter reports an initial attempt to systematize the division of subject matter into program elements. It is assumed that subject matter knowledge consists basically of two classes of statements: rules (generalizations) and examples of rules. These basic subject matter units are taken to be the material from which program frames may be constructed. A matrix procedure is presented for listing and interrelating both the rules and examples of a given subject matter. Methods are described for exhausting the possible fruitful interrelationships between subject matter elements. On the basis of the division of subject matter elements into rules and examples of rules, a variety of frame-types are suggested. Further suggestions are made regarding the sequencing of various frame types starting with the introductory frame and ending with the criterion behavior test frame. Knowledge is taken here to consist of identifiable classes of behavior, behavior which is subject to both strengthening and weakening through environmentally arranged reinforcing contingencies. Chapter 2 also discusses the evaluation and improvement of programs on the basis of student performance and subject matter requirements.

Chapter 3: An Investigation of "Teaching Machine" Variables Using Learning Programs in Symbolic Logic. One of the initial steps in carrying out the objectives of the project was the construction of a program learning sequence designed especially for use in laboratory studies of the effects of programming variables. The subject matter chosen for this "apparatus" program was based upon symbolic logic operations. This subject matter seemed likely to permit the development of a "standard" program, i.e., a learning task which would be relatively free from previous learning influences. Presentation of the subject matter was designed so as to be relatively independent of ordinarily encountered subject matters, similar to the way in which nonsense syllable lists are designed as standard laboratory tasks. The symbolic logic subject matter also presented a manageable unit of learning which would require a relatively short period of study time. A basic program was designed to teach the construction of short deductive proofs involving 15 rules in symbolic logic. The program made no assumption of training beyond the ability to read and to follow instructions. Four experimental variations of the program required different modes of responding from college student subjects. Response modes were: (a) response composition with immediate confirmation, (b) response composition without immediate confirmation, (c) response composition without overt responding, and (d) multiple-choice responding. Two other experimental treatments involved variation in the experimenter's program construction procedure. One version of the symbolic logic program used in these latter two treatments was constructed on the basis of currently available recommendations in the psychological literature. A second version was a program generated by the Ruleg System for program construction. Criterion performance measures consisted of a test requiring short, deductive proofs, a recall test, and a recognition test.

Analysis of the data indicated the following: (a) Experimental variations in mode of responding significantly affected learning time. (b) Error on criterion performance was not significantly affected by mode of responding, including no overt responding at all. (c) Subjects who responded non-overtly to learning the programs took more time on performance tests which immediately followed the program than did subjects who made their responses overtly. Such differences in test time disappeared after a retention period of one week. (d) Differential retention effects were observed as a function of the type of criterion performance measured. A systematically constructed program produced, in less time, criterion performance comparable to that of a less systematically prepared program.

Chapter 4: A Study of the Effects of Partial Reinforcement on Behavior in a Programmed Learning Situation. A major principle of learning describes the effects of supplying reinforcement on a non-continuous schedule; when a response is reinforced on any other than a continuous schedule, the response will tend to attain a higher strength than if it were reinforced on every occurrence. The study reported in this chapter investigated the effects of both continuous and non-continuous schedules of reinforcement on programmed learning. The "standard" symbolic logic program was rewritten and revised for use with high school students. Four groups of high school students took the experimental program, and each group received response confirmation on one of four schedules of reinforcement: 100% continuous reinforcement, 50% fixed ratio reinforcement, 50% variable ratio reinforcement, and 25% variable ratio reinforcement. At the end of the program, all subjects were given a multiple-choice criterion test. On this post-test performance no effect was found that could be attributed to differences in schedule of reinforcement.

A general conclusion which may be drawn from this study and the study in the previous chapter seems to be that reinforcement by confirmation of a response is a crucial event only when response probability is low. In a programmed learning situation, a deliberate attempt is made to evoke the correct answer continually from the subject taking the program and thus, a situation exists in which response probability is high. The effects of intermittent (non-continuous) reinforcement would be more likely to appear in situations where the student experiences considerable difficulty, situations which would not usually appear in programmed learning sequences.

Chapter 5: A Comparison of Three Methods of Programming Subject Material for Auto-Instruction. This chapter describes an attempt to compare three widely different techniques of programming. The suggestions and practices of three learning theorists, Cook, Pressey, and Skinner, were incorporated in three versions of the previously described symbolic logic program. These three versions of the program, based on three rather different theoretical viewpoints, were administered to high school students. Students in the "Cook" group were prompted in giving the correct answers to individual frames by having the correct answer available on the face of each frame. Students taking the "Pressey" program studied the material with no particular overt responding and then were tested on the material at several points during their study by being asked to respond to self-scoring multiple-choice questions. Students in the "Skinner" group took a version of the program consisting of relatively short frames, constructed responses, and response prompting imbedded in the frame material. Immediate and retention tests of deductive-proof problem solving were given to all students. The effect of the experimental treatments was primarily on time to learn and not upon errors of criterion performance. The textual prompting and reinforced responding method advocated by Skinner was found to be the most efficient method of programming the material in terms of learning time.

Chapter 6: The Application of Programmed Learning Techniques to Selected Material in a Professional Curriculum: A Study of Retention Effects. In a graduate course in a school of pharmacy, the subject matter related to 13 vitamins was programmed in programmed textbook format. In the beginning of this experiment, the subject matter for two vitamins was presented by traditional lecture methods and the subject matter relating to two other vitamins of comparable complexity was presented by programmed textbook to the same student. This permitted a comparison of the two methods of instruction. Examinations were administered to determine the degree of the students' mastery of the material.. Students were tested immediately after completing their study of the material and again six weeks later. The results showed significant differences between the programmed and non-programmed presentation. On both immediate recall and retention testing, results were consistently in favor of the programmed presentation of material.. An opinion questionnaire administered to the students at the completion of the study showed a preponderance of favorable opinions expressed by the graduate students in favor of the programmed learning procedure. Additional results of this study indicated that the programming technique lent itself extremely well to instruction at a professional, graduate level. There was a marked decrease in the required teacher time spent in routine lecturing, and the teachers found a greater amount of time available for individual work and for discussion.

Chapter 7: An Investigation of Visual Versus Auditory Programming in Teaching Foreign Language Pronunciation. This chapter describes the development of a rather new technique of response prompting in which visual cues were employed to prompt an oral response. The program used in this study was designed to teach the pronunciation of Spanish letters, syllables, and words. A system for prompting the proper pronunciation of elements of the Spanish language was developed. This system used English language equivalents, in combination with written instructions, to cue an approximately correct pronunciation of the Spanish words. That is to say, transliterations were presented simultaneously with the Spanish words to be pronounced. Three forms of a pronunciation program were administered to students in a high school language laboratory. One version of the program administered to one group of the students was entirely visual and was presented on a succession of index cards. The proper pronunciation of the Spanish language elements was cued by the transliteral prompts. Students spoke the material aloud in the individual booths of the language laboratory. A second form of the program was an auditory version; Spanish words alone were presented on cards and a student-controlled tape recorder was used to prompt correct pronunciation. In this auditory version, students mimicked the pronunciation of a native Spanish speaker in a fashion similar to typical language laboratory procedure. A third group of students used both the auditory prompts on the tape and the visual form of the program. At the completion of the programs, every student read the same list of criterion words into a tape recoreder. Four Spanish teachers and native speakers listened to and judged, on a detailed rating scale, the pronunciation of the criterion words spoken by every student. No statistically significant differences were obtained between the three experimental groups. A result of this study was a demonstration that an approximation of native Spanish pronunciation might be acquired without the assistance of auditory cueing devices. The study also demonstrated a programming procedure which did not employ feedback in the form of a separate response confirmation frame.

Chapter 8: A Vocabulary Program Using "Language Redundancy." While the study reported in Chapter 7 dealt with the teaching of pronunciation, the study described in Chapter 8 deals with the teaching of reading vocabulary in a foreign language. This work represents the development of a new method of programming. The procedure is based on the fact that a verbal utterance contains redundant parts which may be eliminated without loss of information. Redundant parts of a sentence may be replaced by nonsense words without loss in the reader's understanding of the content. If the "nonsense" words are words of a language presently unknown to the reader, then consistent use should equate them to their equivalents in the reader's language. Similarly, the grammatical structure of a new language may be acquired to some degree by phrasing the words of the known language according to the grammatical rules of the new language in so far as possible. In trying out this form of programming, English stories were used in which German words were introduced gradually in place of common English words, the assumption being that the strength of the context would cue the student's understanding of the new German word. New German words were introduced into the text gradually and systematically repeated throughout the course of the test. In later stages of programming, German grammar was introduced.

Details of program construction, format, variations, and tryout results are presented in this chapter. Once again attention must be drawn to the fact that this type of programming, like the programming presented in Chapter 7, does not rely on small frames and consistent response confirmation. In line with the results of the studies presented in Chapters 2 and 3, it was found in the language studies that consistent response confirmation may be relatively unimportant in a case where response probability is high.

Chapter 9: An Exploratory Evaluation of a Discriminative Transfer Learning Program Using Literal Prompts. This chapter reports the results and experiences which came from the development of a small but unique experimental program. The program was based upon and illustrates the application of the discrimination learning paradigm to human learning. In bringing about discriminative control, it is required that a response be elicited in the presence of the stimulus which is to acquire discriminative properties and that reinforcement be contingent upon such an appropriate occurrence. At the beginning of discrimination training, the response, over which control is sought, is assumed to exist at some strength in the repertoire of the organism. If the response is weak or improbable, it can be elicited in the presence of the discriminative stimulus by presenting in addition some stimulus which previously was brought to control the response. Such auxiliary stimuli which are initially used to prompt the response may, as training proceeds, be gradually dropped and the control of the response transferred to the new stimulus. This is the basic paradigm for the programming procedure which is presented in this chapter. Using this procedure with verbal behavior, it is possible to substitute one stimulus for another in cases where both stimuli evoke the same form of behavior. The program was designed to teach first grade children the reading or recognition of eight common color words. The program consisted of individual frames, one per page, placed in a small booklet. Each page of the program contained the word for a common color, and colored lines of the appropriate color were used to prompt the correct response to the word itself. These prompting lines were gradually diminished throughout the program leaving only the word itself. The verbal discriminative response to the colors themselves, learned previously by the child, was transferred to the written words by this means. In essence, the program presented a color itself with a word that stands for the color and gradually removed the stimulus for the old discrimination leaving a new discrimination in its place. Chapter 9 describes the program, the principles of learning on which it was based, the tryout results, and some principles of programming which derive from the use of this "discriminative transfer" paradigm.

Chapter 10: Symbols for Use in the Editing of Programmed Learning Sequences. This chapter represents an attempt to systematize the training of programmers to facilitate the technical aspects of program production. A system of editorial notations which the project psychologists found useful in editing and commenting upon the programming work of beginning programmers is described. It was found that these symbols and notations greatly facilitate the editing process. The symbols presented a means for continually emphasizing the principles underlying good programming and for pointing out pertinent frame characteristics. Use of the symbols served to eliminate the need for time-consuming hand-written editorial notations.

Not Included in This Report: Programmed Clock Reading. An Example in the Analysis of Subject Matter Knowledge. A long standing interest of project personnel was the development of a program for pre-literate first grade children which would be able to teach an ability to tell time from a common clock. This task was undertaken to illustrate certain principles of programming and to provide an example of the analysis of a simple knowledge domain. Throughout the last year of work on the project, the time-telling program underwent at least four major revisions and was tried out in its various forms on 15 children. Basically, a "knowledge" of how to read a clock and tell the time consists of a series of visual discriminations and their associated verbal responses. A child must first learn to discriminate between day and night (A.M. and P.M.). This is necessary since the 12 hours on the clock are repeated twice during each 24-hour period. With respect to the small hour hand on the clock, the child must learn twelve discriminations. With regard to the long minute hand, the child is required to learn 60 discriminations. Such finite limits appear to describe the ability to read a clock; and this, at first glance, is a strikingly simple terminal behavior for a program.

Examination of previous attempts to teach time-telling to children indicates, however, that this is a difficult behavior to achieve. This is perhaps due to the limited amount of behavior which the pre-literate child brings to the programmed learning situation. In other words, programs for very young children carry by far the major burden of stimulus support while programs for older children can assume a great deal of past learning. Previous attempts to teach children time-telling, in the form of books and other devices, indicate a profound lack of specificity in terms of what is to be taught. In the course of the project, the experimenters concerned with the time-telling program had still not reached what they felt was an optimal time-telling program. Their experiences will eventually be written in a further report, but further testing and program development is required. Any present report would be highly tentative and based on inadequate try-out results. One conclusion from the efforts to teach children to tell time in program form is that it is often more difficult to program simple subject matters for very young children than it is to program "highly difficult" subject matters for older children. At this point it seems far more difficult to program time-telling for first graders than it is to specify and program the knowledge of calculus for college students.

Implications

The work carried out by the project points toward certain provocative research problems in programmed learning. These problems can be discussed under the following headings: (1) the analysis of knowledge for converting it into programmed learning sequences, (2) the restructuring of subject materials in order to facilitate the attainment of the objectives of instruction, (3) the nature of learner responses that indicate the attainment of knowledge, (4) the characteristics of reinforcement in a programmed learning sequence, and (5) programs as research apparatus.

1. The Analysis of Knowledge for Converting It into Programmed Learning Sequences. In the construction of programmed learning sequences, a problem arises concerning the way in which a knowledge domain should be analyzed to facilitate the programming enterprise. This concerns the resolution of knowledge into units that can provide the building blocks by which the learner understands the programmed material. There is also the problem of the sequencing and flow of these units. The Ruleg System is primarily concerned with the nature of the blocks themselves, although some characteristics of rule and example sequence frames are considered. In the absence of specific operational procedures for analyzing knowledge for programming, the Ruleg System is an initial crude attempt toward defining such procedures. It is difficult to foresee what form learning sequences for particular subject matters will take when they are systematically analyzed for programming. It is likely, however, that revised knowledge structures and revised subject matter presentations will result from the interaction of subject matter structure and the structure that is suggested from experience in shaping behavior.

2. The Restructuring of Subject Materials to Facilitate the Attainment of the Objectives of Instruction. The work reported on redundancy programming, discriminative transfer programming, and transliteral prompting appears to imply that the ways in which subject matter stimuli are presented to the learner and the ways in which he responds to them may be different from what is the case in conventional subject matter presentation. This work also suggests innovations in the format and style of present concepts of a typical program frame. These programs should remind programmers that techniques are still quite flexible and that much remains to be discovered about programming techniques that carry individuals from initial behavior to subject matter competence. Efforts to program can result in unconventional learning sequences, and a subject matter structure set up for epistomological reasons may be different from the structure required for behavioral reasons. Evidence must be obtained on whether the formal structure of a particular subject matter as developed by subject matter experts is the optimal arrangement for producing creative and problem solving behavior.

3. The Nature of Learner Responses that Indicate the Attainment of Knowledge. This involves primarily the question of response mode--how a learner responds in going through a program and what sort of response is used to indicate terminal knowledge. Criterion performance following programmed learning indicated no significant differences as a result of modes of responding in the studies described in this report and in other studies in the literature. On the basis of these data, one might hypothesize that organisms display "topological generalization," that is, response generalization from the performance of a response in one mode to performance in another mode. For example, it is likely, and indeed evidence suggests, that after going through a multiple-choice discrimination program, a subject can construct his responses when called upon to do so. Learning to make the appropriate discriminations permits him to monitor his own behavior. Thus, it seems that if a program sets up, as a good program does, high probability responses, then the problem of channelling these responses into a number of response modes may be almost trival; this is especially so in an adult subject who, if he can make a response such as "United States Office of Education," will show a high correlation of responses when he is required to write, type, or recognize the response from a list. For low probability responses of the kind that occur in non-programmed learning situations, this topological generalization is probably less likely, and the necessity for response modes similar to the criterion behavior is probably more desirable in the course of instruction.

4. The Characteristics of Reinforcement in a Programmed Learning Sequence. It is possible that the emphasis on providing confirmation and feedback to the student as he goes through a program has led programmers to overlook the notion that much of human learning is "intrinsically" reinforcing. That is, a response often reinforcing consequences, and the student needs only to make a response to provide reinforcing stimuli. In the redundancy program, and in the literal and transliteral discriminative transfer programs, a stimulus was presented to which the student responded, but there was no special frame supplying prompt response feedback. The studies on variations in amount and delay of response showed no significant differences in criterion performance. As a result, it may be hypothesized that in the high probability response context of a program, use can be made of reinforcers that are inherent in responding to the subject matter. In many well-constructed frames, there is little doubt that the student, after having made a response, can check or monitor himself regarding the correctness of his response. In the redundancy and in the color-naming program, the response may be considered "self-reinforcing" perhaps in the same sense that a student is reinforced when he is led to discover something or make what he considers to be an ingenious response to a situation.

The notion here is similar to O. K. Moore's term "autotelic" which he coined to describe the quality of an instructional sequence which becomes an end in itself so that performing such a sequence is "intrinsically" reinforcing. If a frame offers a stimulus which has a high probability of eliciting the desired response, the foundation for an autotelic frame has been laid. Finishing such a frame, that is, giving the correct response, is an end in itself. It is in this sense that a frame which is "peaked" in difficulty (that is, it is just right and neither too hard nor too difficult) can provide a stimulus for an "intrinsically-reinforcing frame"--particularly if the program is tried and tested and especially constructed for one subject. When the program is to be used with a number of subjects, the frame is not quite peaked for the individual subject and hence may require confirming stimuli. Perhaps frames can be made more autotelic if they are written with various degrees of prompting so that a subject can, before making a response, select that frame which most probably assures him of making the correct response. This is somewhat related to the notion of branching and suggests the development of a program in which several alternate frames are prepared for every frame that is written so that students can skip frames until he comes to the one which makes the response "almost" obvious. (Glaser & Schaefer, 1961).

5. Programs as Research Apparatus. The work attempting to specify the characteristics of the units of programmed learning sequences such as the Ruleg System and the proofreading notations for program editing leads to a very general problem. It is apparent that programs can become research apparatus in the same sense that "paired associate" and "serial" word lists have been such. If this is to be the case, two things are needed: (a) a set of terms or a taxonomy for describing the properties of a particular program and the differences between programs, and (b) some standard programs for general research use. At the present time, it is difficult to know just what the results of a study mean when an investigator reports that he has used a "Skinner type" program. The "standard" program described in this report has been an attempt along these lines. The importance of the need for descriptive terms for programs is highlighted when one considers that a major goal of programmed learning is the development of a "theory" of the instructional process that will help produce a specifiable teaching technology. A conceivable way of contributing to this is the development of programmed learning sequences for certain subject matters in which the frames can be identified as to their psychological and pedagogical characteristics. Once such characteristics of an instructional sequence can be specifically identified, than a significant step will have been taken toward increasing the educational psychologist's knowledge of learning and teaching.

In addition to the research activities of the project, a range of dissemination and training activities were carried out. Six interim reports were distributed, and copies of the "standard" symbolic logic program were sent to interested experimentors. Several of the interim reports have been requested for use in graduate courses in psychology and education. In addition to the work described in this report, a paper was prepared by R. Glaser entitled Principles and Problems in the Preparation of Programmed Learning Sequences which has been previously distributed. Also stimulated by the work of the project was a book edited by A. A. Lumsdaine and R. Glaser entitled Teaching Machines and Programmed Learning: A Source Book, published by the Division of Audio-Visual Instruction of the National Education Association.

The University of Pittsburgh supported the work of six experienced public school teachers as programming interns in the Programmed Learning Laboratory. These teachers worked with project personnel in program development and research; the results of this interaction have been described in a report to the Coordinated

Education Center of the University of Pittsburgh: "Auto-Instructional Techniques and the Classroom Teacher," edited by J. I. Taber and R. Glaser.

Considerable professional training was acquired among the personnel associated with the project. The chapters in this volume represent two doctoral dissertations and two masters theses. During the life of the project, in addition to project personnel, advanced graduate students in psychology and education were associated with the project and received research training in the important area of the relationship between the experimental study of learning and its application to educational practice.

PRECEDING PAGE BLANK- NOT FILMED

Chapter 2

The Ruleg System for the Construction of Programmed Verbal Learning Sequences[1]

The Ruleg System for the construction of programmed learning sequences is based on the premise that the verbal subject material which appears in a program can be grouped into two classes of statements which may be called ru's (for "rules") and eg's (for "examples"). Learning to identify ru's and eg's is itself a problem in concept formation and therefore, it may be dealt with as a generalization-and-discrimination problem. That is, an attempt will be made to give a range of both ru's and eg's together, all properly labeled, so as to show the distinction between the two concepts. The two concepts, interestingly, are not mutually exclusive. This "relativity" of ru's and eg's in which the same verbal or symbolic statement may sometimes be both a ru and an eg has so far caused little difficulty. On the contrary, it has proved useful in the construction of programs. Instances of statements which can be both ru's and an eg's will be considered shortly.

No one has yet discovered a completely satisfactory definition of a ru, nor are the authors convinced at this stage that one is needed. The strategy will be to exemplify as widely as possible statements which can be classified as ru's. A ru may be a large number of things. It may be a definition, operational or otherwise. It may be a mathematical formula. It may be an empirical law. It may be a principle, axiom, postulate, or hypothesis from any area of knowledge. But the invariant feature of all ru's is that they are all statements os some generality, from which substitution-instances can be obtained. These substitution-instances are called eg's. An eg may be a large number of things. It may be a description of a physical event.

---

[1] The original paper was prepared by J. E. Evans, R. Glaser, and L. E. Homme and was issued as a separate project report by the University of Pittsburgh in August, 1960.

It may be a theorem or deduction of any sort. It may be a statement of relationship obtaining between specific objects, whether the objects are physical or conceptual. But the invariant feature of all eg's is that they are all statements of some specificity, derived from more generalized ru's.

The clearest instances of ru's and their corresponding eg's come from mathematics. For example, the algebraic "a + b = b + a " is a ru, which summarizes compactly as (infinite) number of substitution-instances, or eg's, such as "7 + 2 = 2 + 7" and "3.4 + 8.6 = 8.6 + 3.4." Again, the statement "Unsupported objects will fall toward the earth" is a ru, while "If I release my pencil it will fall" would be an eg, as would "If I release my book, it will fall." From optics comes the ru that "The angle of incidence is equal to the angle of reflection." To exemplify, one may create eg's such as "If the angle of incidence is 30°, the angle of reflection is equal to 30°." Of course, all of these ru's would need qualification to become acceptable, but the point should be clear that ru's involve some generality, while eg's involve specificity, albeit a relative specificity. That is, "3 + 2 = 2 + 3" is in turn a ru for such eg's as "3 stones + 2 stones = 2 stones + 3 stones" and "3 mammoths + 2 mammoths = 2 mammoths + 3 mammoths." The relationship of the ru-eg concept to the preparation of programmed learning sequences and the steps in the construction of a program are now described.

Step 1. Specify the criterion behavior.

At this step the programmer must attempt to outline, as precisely as possible, both the responses he wants from the student at the end of the program, and the stimuli or cues in the presence of which the student will be expected to make these responses. This is the time when questions such as the following must be asked and answered. If the student is studying statistics, will he have to produce the formula for the standard deviation on his own, or will he always have a book available? Is he being prepared to write a short essay comparing two statistical tests or will he be called upon to take a multiple-choice test at the end of the program? In psychology, will the instructor be satisfied if the student can identify and label examples of regression or spontaneous recovery, or will the student be asked to recommend ways of preventing the occurrence of these phenomena? In general, with what stimulus-support may the student provide himself while the instructor is assessing criterion behavior? Can the student use a book? Another student? The instructor? His notes? A table of integrals? The construction and form of the program will differ radically as a function of the criterion behavior chosen. Now is the time to be both realistic and specific in stating what the objectives of the programmed course are.

Step 2. The specification of subject matter rules.

After the criteria for terminal behavior have been specified, the programmer should proceed to write down all the subject matter rules he can think of. In specifying subject matter rules, the programmer should first work without the support of books, notes, charts or advice. All of these will be available later. For a subject matter expert such "academic" objects usually set off pat behavioral sequences which, although relevant to the task, may preclude an initial free-ranging examination of the field. Once the programmer becomes controlled by stimulus supports, he is apt to neglect rules he might otherwise have thought of. Every rule generated during Step 2 should be written down, preferably on separate cards to permit easy rearrangement.

Step 3. The collection of stimulus support in the form of texts, notes, advice, etc.

The programmer should examine these tools systematically for the ru's which are to be built into the students' repertoire. He should also put these rules on index cards. At times the subject matter may yield only eg's, in which case the programmer may have to induce the rule himself.

Step 4. The preliminary ordering of ru's.

The ru's of a program must be presented in some order, but the conventional order in which they are presented in texts is not necessarily the best order for learning. Writing the ru's on separate cards will facilitate experimentations with ordering. Useful methods for ordering rules include complexity (simpler ru's first), chronology (ordering of ru's in time, as perhaps in a history program), spatiality (ordering ru's in space, as perhaps in a geography program), and dependence on other ru's (in a statistics program where it is necessary to introduce ru's about the mean before ru's about variance are introduced). Care should be taken, at this point, not to omit rules or to leave gaps which the student may not be able to bridge.

Step 5. The ru matrix.

The ru's should now be listed vertically like this:

Ru 1
Ru 2
Ru 3

Then, the same ru's listed horizontally like this:

	Ru 1	Ru 2	Ru 3	. . .
Ru 1				
Ru 2				
Ru 3				
⋮				

The use of a matrix is based on the fact that the hallmark of the expert in a subject matter is his ability to interrelate the concepts in his field. The ru matrix permits a systematic method for examining the ru's of a subject singly, in ordered pairs, triplets, and so on. Ru's can be checked for similarities, differences, possible confusion, or any of a host of possible intraverbal connections. The matrix is also not without heuristic value, in that it may prompt comparisons or relationships not previously considered ("How are the laws of thermodynamics related to the laws of optics?" "How is what we know about bread mold related to bacterial growth?").

The upper-left corner of the matrix is reserved for operators for interrelating the ru's. A very general operator is the relator operator. Consider a matrix like this:

RELATE	Ru 1	Ru 2	Ru 3
Ru 1	1	2	3
Ru 2	4	5	6
Ru 3	7	8	9

Cell #4, for example, would ask "How is ru 2 related to ru 1?" This might require a different handling than cell #2, which asks, "How is ru 1 related to ru 2?" If the order of relationship is of no consequence, symmetrically placed cells, such as 2 and 4, 3 and 7, 6 and 8 are redundant. If ordering does matter, the complete matrix permits examination of every ordered pair. Such ordered pairs may be added to the axes as new operants (row and column entries) to permit expansion of the matrix to ordered triplets, quadruplets, etc. For example, the programmer could first relate pressure to volume in the "perfect" gas equation, then put this relationship on the axes of the matrix, and have a new cell relating pressure and volume to temperature.

The major diagonal relates each rule to itself, and the cells on this diagonal can be reserved for definition cells. Thus it is possible to "relate a ru to itself" by defining the ru in terms of some previous behavior which can safely be assumed to exist at sufficient strength in the student to make the definitions meaningful.

Another operator which might be termed the discriminator operator ("discriminate between") is also useful asking, as it does, for a statement of how two rules are different. This would set the stage for discrimination training.

Step 6. The example operator in the ru matrix and the construction of examples for the program.

This is a critical feature of the Ruleg System. The eg's which are constructed must meet a number of criteria. These are:

1. An adequate number of eg's must be generated. It is mainly through eg's that the student will interact with the subject matter. Eg's must be generated not only to exemplify a ru initially, but to provide later practice and review for that ru.

2. The full spectrum of eg's for a given ru must be considered. This means considering special cases, limiting cases, trivial cases, examples with inadequate information, and examples with redundant information. As a good rule of thumb, the first eg for a given ru should be the simplest possible non-trivial example. For instance, for a student's first example of the computation of a variance, use an N of 2, with simple integer scores which produce a mean which is also an integer. Leading off with a simple non-trivial eg for a ru is the sine qua non of good programming. Any temptation to be complex or

devious should be resisted at this time; there will be opportunity for this when the student's behavior has been adequately strengthened. Complex "test items" come later.

3. As pointed out before, to insure adequate generalization of a ru the attached eg's must be as diverse as possible and still exemplify that ru. Also, to insure adequate discrimination between ru's, eg's must be selected which resemble each other as nearly as possible while still exemplifying their respective ru's.

Step 7. Numbering the cells of the original ru matrix to indicate the order in which the ru's are to be presented and exemplified.

It is often useful to start off "down the diagonal," utilizing the major diagonal (the intersection of each ru with itself) as a definition frame. After that, a decision can be made regarding what ru's should be compared and contrasted, or what pairings of rules should be omitted. The programmer can now number in sequence all cells to be dealt with.

Step 8. Assembling the ru's and eg's into frames.

A frame consists of the stimulus material available at any given moment which demands one or more responses on the part of the student. The student makes his indicated response, and then receives immediate feedback as to the correctness or incorrectness of his response. Frames are constructed by judicious selection and combination of ru's and eg's from the population of ru's and eg's which has been constructed. To provide for responses on the part of the student, words, phrases, numbers, symbols and so on are omitted from certain ru's and certain eg's. Such omissions may be indicated with a line, like this: ______________. For instance, the programmer could give an incomplete ru in this manner: The angle of incidence is equal to the angle of ______________, where "reflection" is to be provided by the student. As an example of an incomplete eg involving numbers, there might be:

$$8 + 3 = 3 + \underline{\qquad\qquad}$$

with "8" being the appropriate response. The choice of the response to be made by the student is obviously an important matter. It often is easier for stylistic reasons to omit certain words from a frame than others; however, it is necessary to consider just what the response required by the student can accomplish toward producing the criterion behavior. Does the response review or strengthen the responses elicited in part frames? Does it help set him up for the next step? Does it "test" what he should have learned previously?

To designate the incomplete ru's and eg's, the symbol "~" (read tilde) can be added over the symbols for rules and examples, like this: $\widetilde{\text{ru}}$ (ru tilde) and $\widetilde{\text{eg}}$ (eg tilde). For terminal situations in which criterion behavior is being called for in the presence of minimal stimulus support, use two tildes over the ru and eg symbols, like this: $\widetilde{\widetilde{\text{ru}}}$, $\widetilde{\widetilde{\text{eg}}}$. An example of a $\widetilde{\widetilde{\text{ru}}}$ would be:

The machine formula for the Pearson-product-moment correlation coefficient is __________.

No other hints or aids would be available to the student. An example of an $\widetilde{\widetilde{\text{eg}}}$ would be:

The variance of the following scores: 3, 7, 4, 2, 9 is _________.

Again, no other ru's or eg's which might serve as prompts must be present, or we would have an $\widetilde{\text{eg}}$, not an $\widetilde{\widetilde{\text{eg}}}$.

Frame types. A list of frame-types described in these terms are presented below. Frames of this kind have been found to be most useful in program construction.

1. Ru + eg + $\widetilde{\text{eg}}$. This frame-type is an excellent method for the economical introduction of a new ru. The ru + eg + $\widetilde{\text{eg}}$ can have the student working an example of a brand new rule on the very first frame in which he is exposed to it. By giving him an explicitly stated ru, and one (or more) carefully chosen eg's before calling for a response by means of an $\widetilde{\text{eg}}$, a powerful prompt is set up which makes an error in responding most unlikely. To illustrate a ru + eg + $\widetilde{\text{eg}}$ frame:

To multiply by 100, move the decimal point two places to the right. (ru)
For example, 100 x 2.843 = 284.3. (eg) Also, 100 x 7.374 = _________. ($\widetilde{\text{eg}}$)

This is a critical phase in the development of a program. It seems best not to introduce more than one ru at a time, or to introduce complex ru's too early, or to ignore consideration of the "simplest possible non-trivial example." The first example in many texts in an effort to be realistic is often much too complicated. The primary consideration is that prerequisite behavior must be adequately strengthened before the student can proceed.

A frame with ru + eg + $\widetilde{eg}$ (in that order) is a most effective method of presenting an initial frame for a new ru. However, other combinations are certainly possible. For example, ru + $\widetilde{eg}$ may be useful, but at least one completely worked-out example may be necessary to permit the working of the first eg with ease and efficiency. Why not eg + $\widetilde{eg}$? This would be an *analogy* frame. The chief characteristic of an analogy is that one must *induce* the ru from the first eg, and then apply it to the $\widetilde{eg}$. Rather than run the risk of having the student induce an incorrect ru, it may be preferable to state the ru for him explicitly. Incidentally, this same philosophy leads to the rejection, in general, of the inductive presentation which might be symbolized as: $eg_1 + eg_{2\ldots}\ eg_n + \widetilde{ru}$. Here a large number of eg's are given and the student is asked to state the ru involved.

After the student can recognize and apply a ru with proficiency, then eg + $\widetilde{eg}$ frames are acceptable. But until that time, it may be often hazardous and slow to "sneak up" on a ru through induction and incidental learning. The student should adopt the expert programmer's carefully chosen statement of a ru rather than use his own halting induction-derived statement. Examples in textbooks are typically too few in number or too restricted to indicate the full range of the rule in question. This makes it possible for a student to induce what is essentially an incorrect rule, but one which happens to fit all the examples present. This constitutes another possible source of danger of the induction process.

After initial introduction to a ru through the ru + eg + $\widetilde{eg}$ formula, it is tempting here to write: ru + eg + eg + $\widetilde{eg}$, to indicate that one *or more* complete examples can be used in the initial frame. Depending on the ru involved, it is often a good idea to provide multiple complete eg's so the student will begin to generalize on the ru and not have the ru attached to only a small number of eg's. A number of variations have proved useful for the subsequent frames. The following frame-types represent only a small number of the possible ru-eg combinations, but they have proved particularly useful in frame construction.

2. *Ru + $\widetilde{ru}$.* Ru's typically contain "technical vocabulary" words and terms. These are words which may be quite new to a student, e.g., "perigee" in astronautics, or they are familiar words with new meanings to be attached, e.g., "population" in statistics. Students often are slow in adding such

words to their active vocabulary. If a technical-vocabulary word is defined in one frame, and then called for after subsequent frames, it has been a common experience that the student will often fail to come up with the word, or will come up with the word only with difficulty. It is better technique to force attention to the technical-vocabulary terms by calling for them explicitly soon after they are introduced. To illustrate:

> The Commutation Law permits the interchange of members connected by a "+" sign. (ru) So if you found a case where two numbers had been switched about a "+" sign you could be sure that the ____________ had been used. ($\widetilde{ru}$)

3. Ru + $\widetilde{eg}$. In later frames in a program stimulus support can be gradually withdrawn by giving only the ru before giving the eg. Likewise the previously mentioned "analogy" frame (eg + $\widetilde{eg}$) represents a weaker prompt for the eg than does ru + eg + $\widetilde{eg}$.

4. Eg + $\widetilde{ru}$. An induction frame which can be used more safely when the student has the ru at some strength.

5. $\widetilde{Ru}_1 + \widetilde{ru}_2$. Such frames can be used to compare and contrast two different ru's as part of the discrimination training involving the ru's. This can obviously be extended to $\widetilde{ru}_1 + \widetilde{ru}_2 + \widetilde{ru}_3$ and so on.

6. $\widetilde{Eg}_1 + \widetilde{eg}_2$. For comparing and contrasting examples of two different ru's, or for demonstrating how two ru's can be applied in turn. Again the extension to $\widetilde{eg}_1 + \widetilde{eg}_2 + \widetilde{eg}_3$, etc., is obvious. (Note that frame-types 5 and 6 would be produced by the matrix technique in the programmer's effort to deal systematically with the important intra-verbal connections which make up the subject-matter.)

7. $\overset{\approx}{Eg}$. The terminal frame of a series in which criterion behavior is performed in the presence of minimal cues. For example, the student might be asked to "Compute an analysis of variance on the following three groups of scores ..." or "Describe the acetylcholine cycle."

8. $\overset{\approx}{Ru}$. Although the eg frame is more basic, in the sense that to respond appropriately to an eg, the student must have command of the ru or ru's involved, on occasion the programmer may wish to call for a statement of a specific ru--to exemplify, "What is the second law of thermodynamics?"

9. $\underline{\overline{\text{Eg}}}$. There is another important concept in the ru-eg framework. This is the concept of the "false example" which is symbolized: $\overline{\text{eg}}$(read "egg bar"). The rationale for the $\overline{\text{eg}}$ is this: In addition to the rich intraverbal associations which the professionally competent person in some fields of learning has, he possesses an additional ability. This consists of his facility in detecting and correcting errors, even though he might never make those errors himself. Such "error-detecting" behavior can be taught like any other behavior, i.e., by giving the student a chance to practice it and be reinforced in performing appropriately. Every field of knowledge has its own peculiar booby-traps and sources of confusion which become painfully well-known to the teacher. Rather than have the student fall into these errors, and rely on uncertain punishment effects to correct his behavior, another technique is to apprise him (the $\overline{\text{eg}}$), and indicate that he is to correct it.

Examination of the protocol following a programmed learning session will help point up the sources of student error, and subsequent programs can incorporate $\overline{\text{eg}}$'s at the appropriate spot. An example of an $\overline{\text{eg}}$:

In his homework a student had the following equation:
$3(5 + 2) = (3 + 5)(3 + 2)$. But you know this is a mistake.
The correct equation is: $3(5 + 2) =$ ____________
Answer $(3 \times 5) + (3 \times 2)$

Another use of the $\overline{\text{eg}}$ is in getting the student to discriminate where and where not to use technical-vocabulary words. For example:

Names of people and places are termed proper nouns. For example, "Chicago" is the name of a place, so it is a proper noun. "Tree" is not the name of a person or place so it is not a ___________.

To summarize the use of the $\overline{\text{eg}}$: When inspection of the ru matrix or program protocols indicate error-likely spots, present the error *labeled as such* and permit the student to deal with the error while still receiving positive reinforcement (confirming feedback).

Step 9. The assembly of frames of various types into a program using the numbered ru matrix as a prompt.

Among the problems to be faced at this step are the following: How many frames should be used for each ru? How should these frames be distributed throughout the program? The procedure recommended here is to lead off with a ru + eg + $\widetilde{\text{eg}}$ frame, and to terminate when the student can deal effectively with $\widetilde{\widetilde{\text{eg}}}$ frames (and $\widetilde{\widetilde{\text{ru}}}$ frames). In between, the general principle is to withdraw stimulus support "gradually," with the use of such frame-types as ru + $\widetilde{\text{eg}}$, eg + $\widetilde{\text{eg}}$, and eg + $\widetilde{\text{ru}}$ as mentioned before. Actually, prior to the first administration of the program, the optimal number and distribution of frame-types within a sequence can only be estimated. This leads to the next step.

Step 10. Tryout and "item analysis" of results.

Only after a frame-by-frame analysis of students' recorded responses to the program can the effectiveness of a program be assessed. If the students are dealing effectively with $\widetilde{\widetilde{\text{eg}}}$'s then the program has done its job and further modifications can be in the direction of making it more efficient without interfering with its effectiveness. If the student is not producing the specified criterion behavior reliably, then steps must be taken to add more frames, redistribute the frames, clear up ambiguities, and so on. One of the chief strengths of the programming technique is its provision for analyzing and revising material on a detailed empirical basis for such analysis and revision, i.e., the responses of the learner.

Step 11. Revise the program on the basis of students' responses and comments.

Students should be encouraged to record their comments and criticisms on particular frames as they proceed through the program. The most difficult phase consists of convincing the student that when he makes an error the blame is on the program and not on himself. This often takes some doing, for most students typically enter the programming situation with a long and consistent history of having their faults and inadequacies pointed out to them. Near-errorless learning may come as something of a shock to them. But the student must be convinced that he is the most important participant in the effort for systematic improvement of the program. After initial administration of a program, a large proportion of the corrections will follow quite obviously from the nature of the errors made. Some will be a result of the simple typographical errors; others will be produced by easily-corrected ambiguities in the text. Failure to come up with a correct term or solution implies the necessity of additional prior strengthening of the concept involved.

To pay lip service to the idea of the constantly revised program is one thing; to do the actual dog-work is another. This is why it is risky to use any elaborate duplicating procedure early in the game. Frames written on separate index cards will probably get added to, removed, and edited. Programs prepared in ditto or mimeographed page form may be discarded and replaced by revised programs. But programs on long continuous tapes or nicely bound and printed are probably "locked in," and more resistant to change. The point is that a sequence is being prepared which will teach a specified behavior, and it should not go to press until this program has demonstrated that it can do its job.

Precautions. In the process of program revision, there are a number of booby-traps. The first is the danger of *over-prompting*. If a student cannot come up with the word "entropy," the solution is *not* to give him "entrop_," and have him guess the letter "y." With adequate prompting, a jungle savage can get through a course in differential equations. It is easy to correct a program by adding prompts, readministering it, and finding to one's delight that the errors drop practically to zero in the corrected frames. But the proof of the pudding is in the eg's; only if the student can produce criterion behavior with minimal cueing has a revision succeeded.

A second booby-trap consists of assuming that the student will respond appropriately to supplemental instructions, comments, and material which appear on the frame but is not involved in the actual production of the response to that frame, for example, "learn this," "read each frame carefully," etc. Such techniques are probably not without effect, but they have a hard-to-control differential effect on students of varying compulsivity. A more realistic rule-of-thumb is: *The student will attend only to those stimuli which are necessary to produce the correct response on any given frame.* Exhortations to learn well and not to make errors are tenuous sources of control of student behaviors; behavior important enough to be strengthened at all should be dealt with by evocation and reinforcement.

The third booby-trap is somewhat allied to booby-trap number two, and is intimately related to the Ruleg System itself. It stems from the fact that in formalizing a system of program writing, as has been attempted in Ruleg, the responses of the student may fall under the control of the wrong set of stimuli. Many people in their childhood have had experience with a "teaching machine" in which the learner used a wire to connect a question, e.g., What is the capital of Arizona?, to possible answers. By joining the correct question and answer, an electrical circuit was made which rang a bell. The difficulty with such a device was that the learner soon learned the relative position between questions and answers, and was able to give the correct answer independently of the verbal material which was on the device. An analogous danger exists in formalizations such as Ruleg. For example, if all technical-vocabulary words are underlined the first time they were presented and always called for later on in the same frame, the production of this word might soon fall under control of the underline, rather than under control of the textual material present.

On the other hand, formalization is not without its benefits. Observational evidence indicates that some sort of a "learning how-to-learn" phenomenon can take place. That is, once a student gets into the swing of the ru + eg + $\widetilde{\text{eg}}$ formula, he begins to grasp new ru's very quickly because he can begin to anticipate how the presentation of the rule will be developed. If he finds that he is always required to state each ru at some later frame (as in a $\widetilde{\text{ru}}$), this may lead him to attend more carefully to the ru as it is first presented.

Step 12. Repeat the administration and revision procedures until the program is producing criterion behavior reliably and efficiently.

This last step concludes the presentation of the Ruleg System of program construction. This early attempt to formalize and make explicit some of the rules of programming will be judged by the same criterion used to ascertain the effectiveness of a program, namely, by results. Like a program, it will be edited, revised, supplemented, or discarded on the basis of the results it produces. In particular the authors would like to dispel any impression that they are irrevocably committed to a Ruleg approach, or that they feel Ruleg will solve all programming problems. The system contains a large number of gaps and guesses. But in programming work it has been found to be an enormously helpful prompting system. Previous to a semi-formalization of Ruleg, the authors' programming behavior could only best be described as sort of an artistic manipulation of the subject matter. Also the Ruleg System provides some way of communicating to others in operational fashion, some of the ways in which program construction is carried out.

Chapter 3

An Investigation of "Teaching Machine" Variables Using Learning Programs in Symbolic Logic[1]

Before programmed learning may be given a fair trial as a teaching method, in comparision to current and traditional methods, a great deal of research will need to be carried out in order that optimal conditions for programmed learning may be specified. Although a number of writers have suggested techniques and practices to be employed in the writing of programs (Gilbert, 1958; Skinner, 1958; Smith, 1959; and others cited in Lumsdaine & Glaser, 1960), these suggestions arise, for the most part, from generalizations from research conducted in contexts other than that of programmed learning. The type of research needed is exemplified in a study by Silberman and Coulson (1959) who compared: (a) multiple-choice versus construction of responses, (b) branching versus no branching, and (c) inclusion versus exclusion of redundant steps in the program. The main conclusion to be drawn from the Silberman and Coulson study was that smaller program steps (frames) are superior to longer steps. The results on the "mode of response" and "branching" variables were inconclusive. Evans, Glaser, and Homme (1960) have presented data in agreement with the Silberman and Coulson finding on the size of step variable and also reported that improved performance resulted from the inclusion of redundant (review) items in a program.

The small experimental literature presently available on program characteristics and efficiency may be summarized as follows: (a) multiple-choice devices which provide immediate knowledge of results can be used effectively to supplement regular classroom instruction (Pressey, 1926; Pressey, 1927; Pressey, 1950; Little, 1934; Briggs, 1947; Jensen, 1949; Jones, 1954); (b) programmed presentation of material, either with or without using a hardware machine, has generally produced better criterion performance than non-programmed presentation even without the refinements which further research will bring (Porter, 1958; Homme & Glaser, 1959; Glaser, Homme, & Evans, 1960); (c) the programming rule to "use large numbers of small steps" seems to be substantially upheld (Silberman & Coulson, 1959; Evans, Glaser, & Homme, 1960); and (d) results of a study of "implicit" responding, responding with no overt spoken or written behavior, casts some doubt on the necessity for an overt response at every step of the program (Evans, Glaser, & Homme, 1960).

---

[1] This is an abridged and edited version of a dissertation submitted by James L. Evans to the Graduate School, Division of Natural Sciences, University of Pittsburgh (April, 1960), in partial fulfillment of the requirements for the degree of Doctor of Philosophy. A detailed version was published as an interim report to the Office of Education: Evans, J.L., Glaser, R., & Homme, L. E., An investigation of "teaching machine" variables using learning programs in symbolic logic, University of Pittsburgh, December, 1960.

One problem in the new field of programmed instruction is the lack of standardized research tools, particularly standardized programs. A standard learning program which would teach some clearly defined verbal or symbolic behavior and would lend itself to experimental variation would facilitate research considerably. Topics drawn from mathematics or logic have much to recommend them. The criterion behavior, as well as the stimuli or cues in the presence of which such behavior is to be produced, can generally be clearly specified. An additional advantage is that a wide range of levels of complexity can be chosen. A complete rationale for the choice of a task in symbolic logic, the topic used in the present study, will be presented in the section on "Materials and Procedures."

Independent variables of possible relevance to programmed learning are numerous, and several writers (Galanter, 1959; Lumsdaine, 1959; Carr, 1959) list suggested variables in some detail. One obvious variable, which we might term "mode of response" is of particular interest, since it forms a point of departure between Pressey and Skinner, the two chief pioneers in the field of automated learning. Pressey (1926; 1927) consistantly used the multiple-choice technique. Skinner (1958) insisted on a "composed" response, which does not involve the added cues and distractions that multiple-choice answers appear to provide. A third mode-of-response variation would be to demand no overt response at all by the S, following pilot studies by Evans, Glaser, and Homme (1960).

A second variable of interest involves the possibility of "formalizing the process of program construction." Several writers (Skinner, 1959; Gilbert, 1958; Smith, 1959) have suggested various techniques and types of program steps or "items." Such suggestions are generally insufficient to instruct inexperienced personnel in program construction. A programming methodology developed concurrently with the present study (Homme & Glaser, 1960; Evans, Homme, & Glaser, 1959; see Chapter 2) gives specific suggestions both for basic types of items and for techniques of assembling these items into a program. Since it is possible to so construct a program in this way "according to formula," such a program would lend itself much more easily to experimental additions, deletions, and reorderings. A demonstration that formally generated programs can compare favorably with programs produced by less specifiable techniques would represent a valuable step in program technology. This would simply be a test of the effect of introducing systematic procedures into program writing, as oppossed to "artful" or intuitive programming.

Finally, most of the proponents of machine learning (e.g., Pressey, 1926; Skinner, 1958) have emphasized the importance of the immediate feedback or confirmation of results which the machine provides. The necessity for such feedback for the most effective modification of many kinds of behavior is well-documented (Estes, 1960). As such, temporal delay of confirmation constitutes another potentially relevant variable. Such delay could be controlled mechanically. In a non-machine presentation, however, delay of confirmation is accomplished by program items which require more than one response. Such items delay confirmation until all responses to that item have been made. Little (1934) demonstrated the importance of the immediate feedback provided by a machine using a non-programmed set of multiple-choice items. Performance measures for groups which received immediate knowledge of results appears to be better than that of groups whose responses were scored and returned the next day. Whether such immediate confirmation of results is critical in the program situation is an experimental question.

The purpose of the present study, arising from the foregoing considerations, was three-fold: (a) to explore the suitability of a task in symbolic logic as a topic to be presented with learning programs of the teaching machine type; (b) to develop a "standard learning program," as well as reliable criterion measures of the material presented in the program, with features which would facilitate further research in the area of programmed learning; and (c) to investigate the effects of variations in methods of responding, program construction, and immediacy of feedback on measures of rate of learning and on immediate and delayed performance measures.

Materials and Procedures

A Rationale for the Use of a Task in Symbolic Logic

A primary purpose of the present investigation was the development of a standard learning program which would provide a uniform laboratory task for other studies in programmed learning. The choice of the topic to be presented in the program will obviously have implications as to the variables which can be studied. This section presents a detailed description of the features of the task selected.

Moore and Anderson (1954) have suggested the use of a symbolic-logic task drawn from that branch of logic known as the "calculus of propositions" for use in studies of human problem solving. Although the task considered in the present study was primarily a learning task rather than a problem-solving task, several advantageous features outlined by Moore and Anderson still hold. The following list describes features of the calculus of propositions (adapted in part from Moore and Anderson) which make such a calculus a particularly appropriate subject matter for an investigation of programmed learning in college students:

1. No assumption of previous training beyond being able to read and follow written instructions needs to be made.
2. Few subjects are likely to have experience with the subject matter.
3. Programs in symbolic logic can be used over a wide range of age and education.
4. Problems of any desired degree of complexity can be generated.
5. Length of program can be shortened or expanded as desired by varying the number of logical rules and concepts which are to be taught.
6. A number of dependent variable measures are available.
7. Learning time appears to fall within practical limits.
8. The task appears intrinsically motivating enough for experimental purposes.
9. Detailed records of subjects behavior both during the program and on criterion measures can be kept.
10. Isomorphic and formal relationships between the calculus of propositions and topics such as the calculus of classes, Boolean algebra, and switching-circuit operations make possible a large number of studies in the area of transfer of training.

In summary, a program designed to teach deductive-proof behavior of the calculus-of-propositions type was selected as the "apparatus" for investigating variables relevant to programmed learning. The calculus of propositions is suggested as a particularly flexible and suitable topic for programming experimentation because of the properties listed above.

The Logical Properties of the Material Presented by the Learning Programs

The behavior which the learning programs used in the present study were designed to produce was: (a) the ability to recognize and employ certain elementary rules taken from the calculus of propositions; and (b) the ability to use them in proper combinations to construct deductive proofs. The importance of the calculus of propositions for deductive systems in general has been outlined by Eves and Newson (1958). This calculus was developed axiomatically in *Principia Mathematica* by Whitehead and Russell (1925). Following this classic contribution to the foundations of both symbolic logic and mathematics have come a number of modifications of the Whitehead-Russell development (see particularly Bernays, 1926; Rosser, 1953; Ambrose & Lazerowitz, 1948). On an empirical basis certain of the theorems and postulates of the calculus of propositions have shown themselves to be particularly useful in the analysis of deductive proofs. Copi (1954) lists 19 "elementary valid argument forms," and the 12 "rules" developed by Moore and Anderson (1954) for problem solving overlap considerably with the set selected by Copi. An even more extensive list of such argument forms is given by Reichenbach (1947). The selection of a particular set of such theorems, rules, or argument forms is of course arbitrary, and is to be judged only by the suitability of the rules chosen for the task at hand. For present purposes a set of 15 rules was selected. These rules involved a total of four different logical operations (negation, conjunction, inclusive disjunction, and implication).

These operations were used on propositions which are symbolized by letters of the alphabet. In the learning programs Ss were not instructed that they were participating in a task involving the calculus of propositions, or that, for example, "$a \rightarrow b$" could be read "a implies b." Rather the task was presented as a purely formal one in which Ss had to learn to manipulate the symbols and letters given in accordance with the rules provided. This procedure seemed warranted to reduce as much as possible transfer effects from verbal habits already present in Ss. That is, if the sequence stated that "$a \vee b$" means "a or b," Ss would have been able to utilize whatever associations they had to statements using "or" to facilitate (or interfere) with their manipulations of "$a \vee b$."

The particular calculus which the learning programs presented had the following features.

1. *Elements.* These were represented by letters of the alphabet (e.g., "a," "g," "k," etc.).

2. *Operators.* These were four in number and included the symbol for negation ("¯"), conjunction ("$\wedge$"), inclusive disjunction ("$\vee$"), and implication ("$\rightarrow$").

3. *Rules.* Examples of the 15 rules are presented in Table 3.1.

4. *Punctuation.* This was provided for by a system of underlining to indicate the specific grouping of associated elements.

An examination of the presentation of the calculus as presented in the programs will reveal several modifications of more conventional procedures.

First, the operators of the system were given specific names to avoid attaching any specific verbal associations to them. The symbol which negated a proposition ("¯") is called "bar," and was used like this: $\bar{a}$. The symbol for conjunction ("$\wedge$") was called "tent," for disjunction ("$\vee$") was called "wedge," and for implication ("$\rightarrow$") was called "spear."

Second, the rules of the system were given novel names, again to control for differential verbal associations by Ss. That is, to call the rule which permits Step 2 here:
1. $a \vee b$
2. $b \vee a$

the "Commutative Law" might immediately identify the rule to Ss who have had algebra, but have little associative value for Ss who had had no algebra or who have forgotten the "Commutative Law." In order to control for this effect the rule was called the "Letter Exchange Rule." An effort was made in the selection of the names of the fifteen rules to select as non-technical and descriptive terms as possible. For example, the rule which permits Step 3 here:
1. a
2. b
3. $a \wedge b$

is called "Conjunction" by some authors (e.g., Copi, 1954) and "Adjunction" by others (Moore and Anderson, 1954). We chose to call the rule the "Tent-Joining Rule" which appeared to provide a more powerful prompt for the actual operation involved. Other classical names for rules provide even poorer prompting cues (e.g., "Modus Ponens," "tertium non datur") and hence were discarded and replaced by simpler and more descriptive terms.

Third, a novel method of punctuation was adopted, again in the hope of simplifying the presentation. It seemed that punctuation using the parenthesis-brace-bracket (e.g., Eves and Newson, 1958), multiple parenthesis (e.g., Copi, 1954), or multiple-square-dot (e.g., Whitehead and Russell, 1925) systems were too difficult to present rapidly and efficiently in a learning program. In particular, the dot notation of Whitehead and Russell becomes quite complex with involved notions of scope attaching to the number of square dots separating propositions, with the added discrimination hazard of using the square dot both as a punctuating device and as an operating symbol (conjunction). The method used in the programs employed underlining to indicate how particular elements are connected by the binary operators (conjunction, disjunction, implication). The scope of the singulary operator (negation) is indicated simply by extending the "bar" over the entire length of the expression to be negated.

Punctuating systems in which the symbols (e.g., "{" or ")") are placed on the same line with the elements and operators appear to present more difficult discrimination problems than does the underlining system. In addition, since the "bar" symbol is also an indicant of scope, corresponding underlines can be dropped when no ambiguity is possible. For example, $\underline{\underline{\overline{a \vee b}} \rightarrow c} \wedge \underline{\overline{e \vee g}}$ becomes $\underline{\overline{a \vee b} \rightarrow c} \wedge \overline{e \vee g}$. Again, the features of the particular calculus of propositions used in the present study were adopted in the hope that such notation would be simpler and would facilitate learning from the program.

Since the primary purpose here was to construct a "piece of apparatus" to be used to investigate programmed verbal learning, the completeness of the rules, from the standpoint of symbolic logic, is in a sense irrelevant. However, it seemed desirable to incorporate rules of as general utility as possible to extend the range of criterion behavior (construction of deductive proofs) which could be demanded. On the other hand, it would have been difficult to teach all of Copi's "elementary valid argument forms" or Reichenbach's more complete listing of rules in a single experimental session. The particular rules which were selected are either included in the forms listed by Copi and Reichenbach, or can be shown to produce such forms in a very few steps.

Table 3.1 presents: (a) examples of the selected set of rules; (b) the names given the rules by Copi (1954) and Moore and Anderson (1954); and (c) the name used in the learning program.

Construction of the Programs Used in This Study

Two basic programs were constructed to teach deductive-proof behavior involving the 15 rules drawn from the calculus of propositions listed in Table 3.1. The first basic program was developed on the basis of the principles of program construction available in the literature at the time of its construction. This program is called the Initial Program. It was used for a series of four experimental variations to be reported.

On the basis of concurrent pilot studies, certain formal principles of program construction were derived which appeared to facilitate the task of programming a subject matter as described in Chapter 2. An additional program was developed employing these formal procedures. The latter is called the Formal Program,[2] and was used in two additional treatments.

Construction of the Initial Program.

It is difficult to specify the method of construction of the first program for the very good reason that at the time of its construction little programming methodology was available. Papers presenting suggestions for programming (Skinner, 1958; Gilbert, 1958; Smith, 1959) were somewhat helpful in suggesting item types, but gave very little help in problems such as number of items, ordering of items, or positioning of review items. A few sample programs were available (e.g., Skinner, 1958), and initial attempts at programming proceeded chiefly by analogy with these prototypes, following the admonition to "proceed by small steps."

---

[2] Submitted as an interim project report entitled "A Symbolic Logic Program for Use in Laboratory Experimentation," to the Office of Education, October, 1960.

Table 3.1

Names and Examples of Fifteen Rules in Symbolic Logic Incorporated in the Learning Programs

Example	Copi	Moore and Anderson	Program
1. p			
2. $p \vee q$	Addition	Addition	Wedge Adding
1. p			
2. q			
3. $p \wedge q$	Conjunction	Adjunction	Tent Joining
1. $p \rightarrow q$			
2. p			
3. q	Modus Ponens	Modus Ponens	Spear Separation
1. $p \wedge q$			
2. p	Simplification	Detachment	Tent Disconnecting
1. $p \rightarrow q$			
2. $q \rightarrow r$			
3. $p \rightarrow r$	Hypothetical Syllogism	Transitivity of Conditional	Double Spear
1. $p \vee q$			
2. $q \vee p$			
or			
1. $p \wedge q$			
2. $q \wedge p$	Commutation	Commutativity	Letter Exchange
1. $p \rightarrow q$			
2. $\bar{q} \rightarrow \bar{p}$	Contraposition	Contraposition	Spear Exchange
1. $p \vee \bar{p}$	Not Given	Not Given	Wedge Bar
1. $p \wedge \underline{q \wedge r}$			
2. $\underline{p \wedge q} \wedge r$			
or			
1. $p \vee \underline{q \vee r}$			
2. $\underline{p \vee q} \vee r$	Association	Association	Line Shift

(Table continued on next page)

Table 3.1 - Continued

Example	Copi	Moore and Anderson	Program
1. $p \vee \underline{q \wedge r}$ 2. $\underline{p \vee q} \wedge \underline{p \vee r}$ or 1. $p \wedge \underline{q \vee r}$ 2. $\underline{p \wedge q} \vee \underline{p \wedge r}$	Distribution	Distribution	Pairing Rule
1. p 2. $\overline{\overline{p}}$ or 1. $\overline{\overline{p}}$ 2. p	Double Negation	Not Given	Double Bar
1. $p \to q$ 2. $\overline{p} \vee q$ or 1. $p \vee q$ 2. $\overline{p} \to q$	Definition of Material Implication	Definition of Conditional	Spear Wedge
1. $p \to q$ 2. $\overline{p \wedge \overline{q}}$ or 1. $p \wedge q$ 2. $\overline{p \to \overline{q}}$	Not Given	Not Given	Spear Tent
1. $p \wedge q$ 2. $\overline{\overline{p} \vee \overline{q}}$ or 1. $p \vee q$ 2. $\overline{\overline{p} \wedge \overline{q}}$	De Morgan's Theorem	De Morgan's Laws	Three Bar
1. $p \vee p$ 2. p	Tautology	Iteration	Wedge Repeat

The actual construction of the Initial Program proceeded as follows. Each of the 15 logical rules was written on a separate index card. The rules were then informally ordered on the basis of simplicity of operation and number of symbols involved. The initial items of the program described the basic logical symbols and their rules of combination. The first rule was then stated, one or more examples were worked, and then some sort of response was required of the S. Typical responses would be to work a new problem, complete an example, or to state some part of a rule. Subsequent items contained either new rules or examples, or review of rules already presented. No systematic review procedure was followed. All 15 logical rules were presented by the forty-third item of the program. Subsequent items in the program gave examples of how several rules in succession could be used to change a set of given symbols into a "winning" or terminal position. Instructions were included for Ss to justify each step taken by giving the initials of the rule used and the step number or numbers to which it was applied. Items containing a number of problems with various combinations of the basic rules were constructed. A total of 72 items made up the Initial Program. Following construction of the Initial Program, each of the 72 items was typed on a separate 5" x 8" index card to form the rough program. On the back of each card was typed the correct response or responses for that item.

A feature of programmed learning emphasized by Skinner (1958) is the critical importance of allowing the behavior of the student to guide subsequent modifications of a given program. Such modifications are facilitated by the fact that the student or S records his response to each item in the program as he proceeds through it. Items on which Ss make errors can then be scrutinized in an attempt to determine the source of the error. In this way ambiguities can be cleared up, unclear typography changed, and additional examples and explanations added to facilitate inadequately strengthened behavior. The process can then be repeated with additional Ss, and subsequent revisions made until the program produces reliable performance at some acceptable level--the version of the Initial Program used in the experiment proper represents a third major revision based on a careful analysis of the responses of approximately 12 pilot Ss. Pilot work was terminated when the Initial Program was producing over 90% correct responses to program items, and completion time on the program was falling within the desired two-hour interval.

The Formal Program. As indicated previously, the second basic experimental program, termed the Formal Program, was constructed in a much more systematic and specifiable manner than the Initial Program. The rationale of the Ruleg System underlying the construction of this program is presented in Chapter 2.

In order to check whether a learning program written "by formula" could produce results comparable to those produced by the Initial Program, the following procedure was employed. The first item which presented a particular rule gave: (a) a verbal description of the operations involved in applying the rule; (b) one or more examples of the rules; and (c) an incomplete example for S to work. The following item gave an incomplete statement of the same rule and required S to give the name of the rule. The third item in the set gave an incomplete example to which the rule must be applied. This completed the set for the first rule. The second logical rule was then dealt with by the same three-item procedure. Each of the 15 rules was subsequently presented in the same manner.

Following this initial series on the rules, Ss were instructed, as in the Initial Program, that they were to transform certain given positions of symbols into a terminal or winning position by successive application of rules. Complete and incomplete examples of this procedure were given. Ss then received a review series, consisting of two items per rule. On the first item an incomplete verbal description of the operation of the rule was given, and S had to supply the term which completed the statement correctly. The following item presented an incomplete example of the same rule. This review procedure was repeated for all 15 rules.

The next series of items gave complete examples of rules being used in pairs to get to the winning position. On the same item, problems were presented in which the same two rules had to be used to reach the winning position. The final series of items presented only the given and winning positions, with instructions to get to the winning position using any rules necessary. During this last series, no completed examples or prompts of any sort were present on the item to aid Ss in constructing proofs. The construction of short deductive proofs with no external stimulus support was the principal criterion behavior which the programs were developed to produce.

In summary, two experimental sequences were developed in an effort to construct a standard learning task for investigating programmed learning. The Initial Program was constructed following the programming principles available at the time of its preparation. The Formal Program was constructed according to a systematic method of program preparation developed in connection with ongoing research in techniques of programming. Both programs were designed to teach the same behavior, i.e., the construction of short deductive proofs, involving 15 rules drawn from symbolic logic. The program task had these features: (a) the topic was novel to most potential Ss; (b) no assumptions were made on previous experience with logic or mathematics; (c) Ss could be brought to a testable level of proficiency in the task in approximately two hours; and (d) experimental variations could be easily introduced.

The Experimental Treatments

With the programs developed, six experimental treatments were set up for study. Four treatments, using the Initial Program, investigated characteristics of the S's response, and can generally be classed as investigations of response mode. The last two experimental treatments, using the Formal Program, were employed to study the effect of variation in technique of program construction, as well as the effect of a provision for review. The influence of these six treatments on measures of learning and retention constituted the major interest of this investigation. The following paragraphs describe the modifications of the two basic programs which provided these variations.

Response Composition (Treatment RC). Under this treatment, Ss were required to compose their answers to each item. This is the method of responding recommended by Skinner (1958) and used by him in his machine work. No answer of any sort was available on the front of the item. Ss were required to respond by supplying missing terms, working problems, and answering questions. In the following item from the Initial Program, the task of the S was to provide the two answers indicated by the blanks.

Example of a Response Composition Item

These three signs are called connectors: ∨ , →, ∧ .
Each connector has a special name to help us remember it.
This connector is called 'wedge': ∨ .
This connector is called 'tent': ______
This ______ is called 'spear': → .

After the S had read the item, and had recorded the two answers which he considered to be correct, he turned to the back of the index card which contained the item, and compared his response with the correct answers ' ∧ ' and 'connector'. If his responses matched the correct answers on the back, he proceeded to the next item. If he made an error, he was instructed to circle the error on his answer pad and determine why his answer was incorrect before proceeding to the next item.

Multiple Choice Response (Treatment MC). The method of responding utilized in the work of Pressey (1926; 1927) involved selecting the correct answer from a number of alternative answers. Such a method of responding was used in Treatment MC.

In this treatment Ss were given items identical with those presented in RC. However, at the bottom of each item was a set of answers lettered "A," "B," "C," and so on. The task in this condition was to select the correct answers from this set by writing down the letters corresponding to the proper response. The following example shows a multiple-choice version of the item in previous example.

Example of a Multiple-Choice Item

These three signs are called connectors: ∨ , →, ∧ .
Each connector has a special name to help us remember it.
This connector is called 'wedge': ∨ .
This connector is called 'tent': ______
This ______ is called 'spear': → .

A: ∨   B: ∧   C: symbol   D: connector

Ss in this treatment followed basically the same procedure as Ss in Treatment RC, but in this condition they had to choose their responses and record the corresponding letters. They then checked their responses against the letters representing the correct answers and proceeded as before. To the extent that it was possible, the alternate incorrect answers were selected from errors made on the same items by pilot Ss. This provided an empirical method for the construction of false alternatives.

Implicit Response (Treatment IR). The major aspect of this treatment was that Ss did not make overt written responses to the items in the Initial Program. Pilot work reported by Evans, Glaser, and Homme (1960) showed no significant differences in criterion performance between Ss who recorded their composed responses and Ss who confirmed their "implicit" responses, but did not record them. Since overt responding by the S characterizes the work of other investigators in this area (e.g., Pressey, 1950; Skinner, 1958), this variable appeared to deserve further study.

The program for this treatment was constructed by giving the correct answer or answers at the bottom of the item requiring the response. Hence Ss had the correct responses available at all times as they studied the items. Their instructions were to study the card until they understood why the answer provided was correct in each case, and then to proceed to the next item. They were specifically instructed not to write down their answers in any form. An example of an implicit-response item is as follows.

An Example of an Implicit Response Item

These three signs are called connectors: ∨ , →, ∧ .

Each connector has a special name to help us remember it.

This connector is called 'wedge': ∨ .

This connector is called 'tent': ________

This ________ is called 'spear': → .

Answer: ∧

connector

An alternative method would have been to insert the correct answers into context in the sentences and examples of each item. However, the technique of placing the correct answers at the bottom of the card was selected, in order to force attention to the same portions of each item as did Treatments RC and MC.

Immediate Feedback (Treatment IF). Most items in the Initial Program required more than one response. Ss were not allowed to check their responses under RC and MC until they had completed all responses to that particular item. This effectively delayed the feedback as to the correctness of individual responses. Even the last response to an item was not always confirmed immediately. Ss usually checked off their responses with those on the back of the card in order, and such checking procedure delayed the confirmation of the last response. This delay between response and response confirmation under Treatments RC and MC averaged about two minutes, with a range of 30 seconds to five minutes, depending on item difficulty. To determine whether such delay of confirmation influenced performance, the following procedure was devised. A numbered list of answers to all items in the Response Composition form of the Initial Program was prepared. This list, with space provided for responses, was given to Ss along with a small cardboard mask.[3] Ss were instructed to cover the answers with the mask until they had written the first response to an item. At that time they were to move the mask down until the correct answer appeared on the answer sheet. They checked this response, and then repeated the procedure for subsequent responses to that item. In this manner the confirmation of a response was given immediately, regardless of the total number of responses on that particular item.

The four experimental treatments described above employed the Initial Program. In summary, it can be pointed out that all four involved variations in the mode in which Ss responded. Under RC, Ss constructed their responses to each item in full. These responses were then checked with the answers on the back of the card. Under MC, Ss used a choice response, selecting their answers from a set of multiple-choice answers appearing at the bottom of the item. They then listed the letters corresponding to the answers they considered to be correct. Under IR, no overt written response was required. The correct answers, separated from the context as in RC and MC, appeared at the bottom of the item for S to check. Under IF, a masking technique was used to reveal immediately the correct answer following each response to a response-composition time.

---

[3] The masking procedure is a modification of a technique used by Ferster and Sapon (1958) in teaching German composition by a learning program.

Formal Program (Treatment FP). The essence of this treatment was the use of the Formal Program previously described, in comparison with the treatments using the Initial Program. A demonstration that such a program was as effective in producing criterion behavior as previous methods would facilitate operational specifications of program construction.

This treatment was administered exactly like Treatment RC. Responses to each item were composed in full and then checked against the answers on the back of the card. In contrast with the Initial Program, in which items typically required more than one response, most items in the Formal Program called for a single response, until the more complex responses toward the end of the program.

The following example illustrates a typical item of the Formal Program. The rule-example-incomplete example pattern used throughout the program is evident in this item.

Example of an Item from the Formal Program

This sign is named 'wedge': $\vee$.

It is called a connector, since it connects any two letters when it appears between them.

For example, we would write "m wedge r" like this: $m \vee r$

Now you write "k wedge t":

The correct answer "$k \vee t$" would appear on the back of the card on which this item appeared.

Formal Program + Review Card (Treatment FP+). Programming procedures developed thus far had no provisions for "memory storage" of the materials presented by the program. Most programs in existence do not have such a provision. That is, at the time an S is given a review item on materials previously covered, he typically is without any summary, outline, or abstract of such material to prompt his response. The possible relevance of such a factor was emphasized by responses to a questionnaire administered during the early phase of experimentation.

Several Ss indicated that it would have been helpful to have some method of reviewing the rules presented in the program. The following procedure was devised to check the effect of a provision for review on performance.

A complete list of examples of all rules was prepared, and all such examples were typed on a single 9" x 12" card. Ss were instructed that this list of rules would be available for their use as they proceeded through the Formal Program, but that it would not be available during the tests which followed the learning program. No other suggestions for the use of this review card were made. In all other respects the administration of this condition, called Treatment FP+, was the same as in Treatment FP.

In summary, Treatments FP and FP+ involved a response-composition mode on the Formal Program. The two treatments were identical except that under FP+ a review card with a complete list of the 15 logical rules was available as "memory storage" during the course of the program.

Construction of the Criterion Measures

In an investigation of program versus textbook presentation of the same material, Evans, Glaser, and Homme (1960) found that Ss using the learning program performed approximately the same on a multiple-choice test as Ss learning from a test. However, the program group did appreciably better on a completion-type test. This suggests that different experimental treatments may produce differential performance as a function of the criterion measure used. This finding, coupled with the exploratory nature of the present study, suggested that a variety of types of criterion tests should be constructed to sample various aspects of the behavior produced by the learning programs. One "level" of performance would involve making discriminations between correct and incorrect instances of applications of the logical rules. A true-false test was constructed for this purpose. A second type of behavior involves recall and application of each of these rules when the name of the rule is given. A third type of behavior consists of successive applications of these rules in combination to produce short deductive proofs.

In addition to assessing the effects of experimental treatments on different types of criterion performance, it also seemed important to assess treatment effect on retention of the behavior learned in the program. The systematic nature of the chosen task facilitated the construction of parallel retention tests for each of the criterion measures.

The True-False Tests. The first criterion measure was a 15-item true-false test on each of the logical rules presented by the programs. A table of random numbers was used to determine whether a true or a false example of a particular rule would be constructed. An example of each rule was then prepared in which the last step of the example either followed from the rule in question or contained some error. The 15-item test so constructed was designated "TF1."

A parallel, but not identical, retention test was prepared in the following manner. If a true example of a given rule had been given in TF1, a false example of that rule was presented in the retention test, designated "TF2." Also, false examples in TF1 were replaced by true examples of those rules in TF2. In this way each S had to discriminate one true and one false example of each of the 15 rules in the course of the criterion testing.

The Recall Tests. A second type of behavior of interest was the recall and application of each rule given the name of the rule and a step or steps on which to apply it. A 15-item test was constructed by calling for each of the rules in turn. This recall test was designated "R1." By changing the particular symbols involved, a parallel, but not identical, test was constructed for retention purposes. This test was designated "R2."

The Deductive-Proof Tests. The criterion behavior which the programs were primarily designed to produce was that of using the logical rules in combinations to obtain deductive proofs. The following systematic procedure was used to construct 15 problems of this type. First, a 15 x 15 matrix was constructed with a list of the names of each rule forming the axes. Each of the 225 cells of the matrix then represented an ordered pair of rules. By selecting the necessary initial steps, and then by applying two rules in succession, a terminal position deducible from the initial position was reached. To present this as a problem, the initial positions would be given, along with the terminal or "winning" position. Each S then had to provide the intermediate steps which constituted the proof.

Selection of the pairs of rules from the matrix was done as follows. Cells were chosen at random. However, a constraint was imposed to prevent two cells from the same row or same column from being selected. Also excluded were cells along the major diagonal which represented the intersection of each rule with itself. This procedure resulted in a set of 15 problems with the following properties: (a) each rule was employed in two and only two different problems; and (b) each rule appeared in the first position in one problem and in the second position in some other problem. A set of 15 rules so generated was designated "DP1." A parallel retention test, DP2, was constructed using the same matrix and procedure. The only additional constraint was that no cell be used to generate a problem which had already been used in DP1.

The Attitude Questionnaire. A short questionnaire was constructed to assess the reactions of Ss to the method of programmed presentation of material to be learned. Ss were asked to rate their interest in taking a course using programmed material, attitude toward the effectiveness of such procedures, and opinion of the amount of review which the program provided. Other comments on the experiment itself as well as on programmed presentation were encouraged.

Other Experimental Materials

Each S in Treatments RC, MC, FP, and FP+ received a 3" x 5" answer pad on which to record his answers. These Ss were instructed to turn to a new sheet on the answer pad when they turned to each new item in the series. This procedure was adopted to prevent previous composed responses from serving as prompts for subsequent response to any particular item. The answer sheet and mask for Treatment IF have been described previously. Ss in Treatment IR, who were not required to write down their answers, used a 5" x 8" pad to record their rate of responding as described in the next section.

Procedures and Subjects

All Ss in the present study were University of Pittsburgh undergraduate students. Ss who wished to participate filled out contact forms which informed them of the general nature of the experiment and the rate of pay ($4.00 for completion of both experimental and retention sessions). On the average, three Ss were run per experimental session.

Ss were assigned randomly to the experimental treatments. At the beginning of each experimental session, E gave the following instructions verbally:

> Today you will be participating in an experiment in which we are investigating some new ways of learning written material. In front of each of you is a pack of cards. When the experiment begins, you will be reading each of those cards in turn. Each card will indicate that you are to make some sort of a response to the material on that card. Following your response you will always find out whether you are correct or not. You will get one six-minute break about half-way through the cards. You will get another six-minute break when you have

> finished the cards. After that break, you will take three different tests over the material you have learned from the cards. Finally, you will fill out a short questionnaire on your reaction to the experiment.
>
> Now read your instruction cards on top of the pack in front of you. When you have read and understand these instructions, look up at me. When everyone has read the instructions, we will all start through the cards together.

At this point E allowed Ss to read the instruction cards which described their particular procedures. Some information on the instruction cards duplicated the verbal instructions by E. When all Ss had read their instructions and any questions were answered, E said:

> We are about to begin. Go through your cards at your own most confortable study pace, just as if you were preparing for an exam. Some people will finish before others because of different procedures. Do not worry if you seem to be finishing much faster or much slower than others, since they have different tasks to do.
>
> When I say "Write down your card numbers, please," jot down the number of the card you are working on at that time, and go on. Once more, make sure that you work at your own most comfortable rate. If you have a question at any time, raise your hand and I will come to help you. You may begin work.

At three-minute intervals following the start of the session, E would say, "Write down your card numbers, please." This procedure provided a method for getting a rate-of-response record, i.e., number of cards per three-minute interval. During the learning sessions, Ss were placed so that E could observe easily all phases of their reading and responding. This was done to control possible variations from the experimental procedure such as looking at the back of the card before answering or moving the mask before responding under Treatment IF.

Ss took their two six-minute breaks outside the experimental room. The combination of different treatments and different working speeds on the part of Ss resulted in a natural staggering of break time, so that two Ss rarely had their break together. This appeared to be a desirable feature for preventing Ss from discussing their different treatments during the break. After the second break following completion of a program, Ss returned to the experimental room to take three performance tests over the programmed material. Following completion of each individual test, E recorded the time, removed that test, and brought the next test. This procedure was adopted to control for Ss using the results on one test to prompt themselves on another test. After completion of the third test (DP1) Ss were given the attitude questionnaire. Following this, an appointment was made with each S for a retention test one week later. In the retention phase the three retention tests were administered in the same manner as the post-program tests. Time scores on these tests were again recorded.

Results

This section, which presents the analyses of the results obtained in the study, will be developed in the following manner.

First, time scores, both on the learning programs and on the immediate and retention performance measures, will be presented. Second, error scores on the learning program and on the performance measures will be analyzed. Third, properties of the criterion measures such as reliability and range of performance will be presented. Fourth, the influence on performance of individual characteristics such as sex, mathematical experience, and college class will be considered. The final analysis will be concerned with responses by Ss to the attitude questionnaire.

Analysis of Time Scores

The following time scores were available for analysis: (a) total time to complete the learning program; (b) times spent on each of the three immediate performance tests; and (c) times spent on each of the three parallel retention tests. Program times will first be analyzed, and then the immediate and retention performance times will be treated together.

Program Times. A record of total time in minutes which each S spent on the learning program was made. Table 3.2 presents the means and standard deviations of these scores for each of the six experimental treatments. The means have been ordered to facilitate comparisons.

Table 3.2

Means and Standard Deviations, in Minutes, on Total Learning Time Spent on Programs (Treatments have been ordered by mean time)

Experimental treatment	Mean	S.D.
Implicit Response (IR)	81.8	16.9
Formal Program (FP)	103.6	22.1
Formal Program + Review Card (FP+)	104.1	22.2
Multiple Choice (MC)	121.6	23.7
Immediate Feedback (IF)	127.1	19.6
Response Composition (RC)	132.1	17.4

Ss in Treatment IR(no overt responses to the Initial Program) took appreciably less completion time than Ss in the other treatments. The IR group averaged over 20 minutes less learning time than the two groups who composed their responses to the Formal Program (FP and FP+). Mean learning time for these two treatments was practically identical. The Three Initial Program treatments requiring multiple-choice and composed responses (MC, IF, and RC) had mean times from about 15 to 25 minutes more than the means of the two Formal Program groups. Considering the four Initial Program treatments together, it appears that requiring overt responses by Ss increases the mean learning time from 40 to 50 minutes as compared with implicit responding.

Table 3.3

Analysis of Variance of Scores of Total Learning Time Spent on the Programs for Each of Six Experimental Treatments

Source	df	Mean Square	F	P
Total	59			
Treatments	5	3537.44	8.52	<.01
Error	54	414.98		

An analysis of variance of learning times under the six experimental treatments is presented in Table 3.3. The differences between treatment means is highly significant (p. < .01). Two further analyses, now shown, compared the four Initial Program treatments by themselves (treatments significant at the .01 level) and the two Formal Program treatments separately (treatments variable was non-significant). Since the Initial Program and the Formal Program differed in their method of construction and in the total number of items, two further analyses, not shown here, were made.[4] An analysis of variance for the four Initial Program treatments considered separately showed significant differences between means (p. < .01), due chiefly to the distance of the mean of the implicit-response group from the means of the three overt-responses groups. However, the two Formal Program groups (FP and FP+), had essentially the same mean learning time.

In summary, significant differences on total learning times were found due to the six experimental treatments investigated. Consideration of the Initial Program treatments separately indicated that implicit responding to items takes considerably less time than treatments requiring overt composed or multiple-choice responses. Differences between these four mode-of-response treatments were highly significant. As for the two Formal Program treatments, the use of a review card in Treatment FP+ appeared to have little effect on total learning time.

Immediate and Retention Performance Test Times. Time scores were available on each of the three performance tests administered immediately after the learning session. Analagous scores were available on the three parallel forms of these tests administered one week later. The means of the immediate and retention test times, separated by experimental treatments, is presented in Table 3.4. Means of the sums of the three immediate test times, and of the three retention test times, are also presented. Finally, the means of the total time taken on all immediate and retention tests summed together is given.

---

[4] See the original report referred to earlier (Evans, Glaser, & Homme, 1960)

Table 3.4

Mean Time in Minutes on Each of Six Performance Tests for Each of Six Experimental Treatments

Test	Treatment					
	RC	MC	IR	IF	FP	FP+
TF1	5.0	5.0	9.0	6.5	5.1	5.4
TF2	4.2	4.3	4.4	4.5	5.3	4.4
R1	5.6	6.2	7.5	5.6	4.7	4.5
R2	4.4	4.3	3.9	5.0	4.2	3.6
DP1	21.5	20.5	31.2	28.6	20.3	20.4
DP2	21.2	19.9	19.8	24.2	18.3	20.3
TF1 + R1 + DP1	32.1	31.7	47.7	40.7	30.1	30.3
TF2 + R2 + DP2	29.8	28.5	28.1	33.9	27.8	28.3
Total Time	61.9	60.2	75.8	74.6	57.9	58.6

Total Test -- By treating the immediate and retention tests as separate trials, a repeated-measures analysis of variance (Edwards, 1956) can be performed on the time scores. Such a design permits three sources of variation to be tested for significance: (a) differences due to experimental treatments; (b) differences between trials (immediate performance versus retention); and (c) interaction of treatments and trials. The first set of scores to be so analyzed were the total immediate time scores (TF1 + R1 + DP1) and the total retention time scores (TF2 + R2 + DP2). This analysis is presented in Table 3.5. It will be noted that all three sources of variation show significant effects.

Table 3.5

Analysis of Variance of Performance Times under Six Experimental Treatments using Total Immediate and Retention Tests as Separate Trials

Source	df	Mean Square	F	p
Between treatments	5	332.65	2.40	< .05
Between Ss in same group	54	138.35		
Total between Ss	59			
Between trials	1	1092.03	19.93	< .01
Interaction: trials x treatments	5	236.81	4.32	< .01
Interaction: pooled Ss x trials	54	54.79		
Total within Ss	60			
Total	119			

Considering the effects of the experimental treatments, it is interesting to note that the implicit-response condition (IR), which was completed in markedly less learning time, had the highest mean time for completion on the three immediate performance tests. This mean was seven minutes more than that of the next treatment (Immediate Feedback) and over 15 minutes more than those of the remaining four treatments. Such a difference disappeared over the retention interval; the mean total times on the retention tests were quite similar for the experimental treatments. All six treatments exhibited reductions in mean total test time over the retention interval, and this effect is statistically significant. The significant interaction effect appears to be due chiefly to the fact that the mean total test time for Treatment IR dropped almost 20 minutes over the retention interval, while the mean times for the other five treatments dropped only two to six minutes.

True-False Tests -- To explore such findings in more detail, the same method of analysis was applied to the scores of the three types of performance tests considered separately. The results of these separate analyses are not shown in detail in this report, but have been reported elsewhere (Evans, Glaser, & Homme, 1960). The results of the analysis of true-false scores was essentially the same as for the total time scores on the three tests together. All three sources of variation were again significant. The mean time for the implicit-response group again was considerably higher than that of the other treatments on the test taken immediately after the learning session. As before, the mean of this group dropped markedly by the time of the retention test, and became indistinguishable in size from the means of the other treatments. With one exception, all treatment mean times were less on the retention tests, and this effect was significant. Again, it appeared to be the differential drop in mean time for the implicit-response group which produced the significant trials x treatments interaction effect.

Recall Tests -- In the repeated-measures analysis of time scores on the recall tests (R1 and R2), trial effects and trial x treatment interaction was again significant, but a treatment effect was absent in this case. Examination of mean scores on the recall tests in Table 3.4 reveals a reduction in mean time on the retention test for all treatments, with Treatment IR again showing the highest immediate test time and the greatest drop over the retention interval.

Deductive-Proof Tests -- The results of a similar analysis applied to time scores on the deductive-proof tests (DP1 and DP2) showed that neither the treatment effect nor the trial x treatment interaction effect reached statistical significance in this case. However, all treatments showed a significant drop in mean test time over the retention interval, as in the previous analyses. The mean of the IF group was highest on the immediate performance test, as before, and again dropped to the level reached by the other treatments on the retention tests. The drop for the IR group was over 11 minutes. One of the remaining groups (IF) dropped over four minutes, while the remaining four dropped two minutes or less in mean time. Despite these differential reductions in mean test times over the retention interval, the trial x treatment interaction effect failed to reach statistical significance ($.05 < p < .10$).

Since significant differences due to treatments were present in the analysis of mean total time (Table 3.5), and three of the four trial x treatment interactions so far considered were also significant, the following analysis was performed to gain more information about treatment effects. Rather than summing together the immediate and retention time scores as done in the repeated-measures analysis, each of the three immediate tests and each of the three retention tests were considered separately. A one-way analysis of variance was performed on each of these six time scores. The results of the analyses are presented in Table .6. Inspection of this table reveals that significant treatment effects were found on the immediate true-false and the immediate deductive-proof tests, but not on the recall tests. All such differences on time scores due to treatment effects had disappeared by the time of the retention tests.

Table 3.6

Summary of Variance Analyses of Times Spent on
Three Immediate and Three Retention Performance Tests

Score	F (for 5 and 54 df)	p
TF1	4.50	< .01
TF2	0.74	NS
R1	1.56	NS
R2	1.23	NS
DP1	3.36	< .05
DP2	0.88	NS

A summary of the results of analyses of time scores on performance tests is as follows. Mean time scores in general showed consistent and significant reductions between the immediate and the retention forms of the various criterion tests. Experimental treatments produced significant effects on the immediate true-false test times and the immediate deductive-proof test times, with Treatment IR consistently showing highest mean times. No significant treatment effects were found on the immediate recall test.

All such effects attributable to treatments disappeared over the retention interval. Several significant trial x treatment interaction effects were noted in the repeated-measures analyses. These differences appear to be due to the marked drop in mean time over the retention interval which characterized Treatment IR.

Analysis of Error Scores

Error scores, like time scores, were available both on the responses made during the learning program and responses made during the immediate and retention performance tests. The results of analyses of these scores is presented in the next two sections.

Program Errors. In all treatments except the implicit-response (IR) condition, Ss recorded their responses to each item as they proceeded through the program. Such responses were scored as correct or incorrect. A summary of these error scores is presented in Table 3.7. It will be recalled that the Initial Program consisted of 72 items, and the Formal Program consisted of 125 items. However, the Initial Program contained more items calling for more than one response than did the Formal Program. In view of this, it appeared that the total number of responses represented matters more accurately than the total number of items. A count of total responses required in each program was made. The Initial Program called for 189 responses; the Formal Program called for 151 responses. These figures were used to compute "per cent errors per response" in Table 3.7.

Table 3.7

Mean Errors and Mean Per Cent Errors per Response on the Learning Programs for Five Experimental Treatments[a]

Mean	Treatment				
	Initial Program[b]			Formal Program[c]	
	RC	MC	IF	FP	FP+
Errors	34.3	18.1	30.7	25.9	17.1
Per cent errors per response	18.1	9.5	16.2	17.1	11.3

[a] Errors under Treatment IR are not available since Ss did not record their responses.

[b] 189 responses in 72 items.

[c] 151 responses in 125 items.

Inspection of Table 3.6 reveals that almost twice as many errors were made under Treatment RC as under Treatment MC. This appears to be in line with a common finding that is more difficult to construct a response correctly than it is to recognize such a response (e.g., Luh, 1922). It does not appear that providing immediate confirmation of individual responses (Treatment IF) reduces to any marked degree the total learning errors as compared with a delayed-confirmation treatment (RC).

For the two Formal Program treatments (FP and FP+), provision of a "memory storage" device in the form of a review card appears to decrease learning errors somewhat. This seems reasonable since such a review card should provide Ss with an additional prompt to increase the probability of correct responding to review items in the program.

Treatment RC and Treatment FP both involved the same sort of response composition, although each employed a different form of the experimental program. Per cent errors per response is approximately the same for the two treatments. It appears that construction of programs by formal techniques (Formal Program) produces approximately the same percentage of learning errors as produced by the Initial Program.

In summary, composition of responses in Treatment RC produced almost twice as many learning errors as did making multiple-choice responses in Treatment MC. No particular reduction in number of errors was made by providing immediate feedback for composed responses under Treatment IF. Provision of a review card to prompt responses to the Formal Program reduced the number of learning errors. Error rate for composed responses was about the same for Initial and Formal Programs.

Immediate and Retention Error Scores. Each of the three immediate performance tests and each of the three parallel retention tests contained 15 items, making a total of 90 items for each S. Upon completion of the experiment, each of the items was scored, and the errors on each test were totaled and used as the index of performance on that test. Table 3.8 presents the mean error scores for each of the six treatment groups on each of the six performance tests. Means of the total number of errors for the three immediate and three retention tests are also presented, as well as mean total error for all six performance tests combined.

Table 3.8

Mean Error Scores on Each of Six Performance Tests for Each of the Six Experimental Treatments

Score	Treatment					
	RC	MC	IR	IF	FP	FP+
TF1	5.4	5.8	6.5	6.2	6.0	5.2
TF2	3.4	3.7	3.7	3.7	3.8	4.9
R1	3.2	4.9	3.9	4.1	4.6	4.0
R2	4.2	5.6	4.0	4.8	6.2	3.9
DP1	8.1	9.9	7.8	7.9	7.5	6.5
DP2	7.7	9.6	7.7	7.8	8.2	7.1
TF1 + R1 + DP1	16.7	20.6	18.2	18.2	18.1	15.7
TF2 + R2 + DP2	15.3	18.9	15.4	16.3	18.2	15.9
Total Errors	32.0	39.5	33.6	34.5	36.3	31.6

The method of analysis of error scores proceeded in the same manner as the analysis of time scores. The immediate and retention tests were treated as separate trials, and a repeated-measures analysis of variance of both total scores and scores made on the three types of performance tests separately was made.

Table 3.9 presents an analysis of total immediate and retention error scores. The only effect to reach statistical significance was that due to trials. It is interesting to note that the total number of errors made after a one-week retention interval (1000 errors) was less than the total number of errors made immediately after the learning session (1075 errors).

Table 3.9

Analysis of Variance of Error Scores under Six Experimental Treatments using Total Immediate and Total Retention Tests as Separate Trials

Source	df	Mean Square	F	p
Between treatments	5	43.67	-	NS
Between Ss in same group	54	112.92		
Total between Ss	59			
Between trials	1	46.88	7.51	<.01
Interaction: trials x treatments	5	6.97	1.11	NS
Interaction: pooled Ss x trials	54	6.24		
Total within Ss	60			
Total	119			

With respect to other effects, it appears that experimental treatments had little effect on error score, either as a primary source of variation or in combination with trials as a trial x treatment interaction. In light of the marked effect of treatments on time scores, this was a most surprising result. A discussion of these findings taken together is presented in the discussion section.

True-False Test -- Further analyses were made of error scores by separating the true-false, recall, and deductive-proof scores and will be summarized here. Analysis of error scores from the true-false test (TF1 and TF2) gave results essentially the same as in the previous analysis. Again the only main effect to reach significance is that attributable to trials. As before, the number of errors made after the retention interval (232) was significantly less than the number made on the immediate test (351). Effects due to treatments and interaction effects were negligible.

Recall Test -- In the analysis of the recall tests (R1 and R2) the only source of variation to produce a significant p-value was that due to trials. Here, however, the previous finding was reversed. Significantly more errors (287) were made after the retention interval than at the time of the immediate performance test (247). This effect was opposite in direction from that of the true-false and total scores, where significantly more errors were made before the retention interval. A discussion of this differential performance after the retention interval is presented in the discussion section.

Deductive-Proof Test -- The last analysis was that of the deductive-proof scores (DP1 and DP2). Here, no significant differences due to any of the testable sources of variation were found. Treatment and trial x treatment effects were absent as before. For the first time, however, no effect due to trials was found. The total number of immediate errors (477) on the deductive-proof tests was virtually the same as the total number made on the retention test (481).

In summary, it is interesting to note that three types of performance tests were employed, and each type produced different results over the retention interval. The true-false test showed significantly fewer errors after one week; the recall test showed significantly more errors after one week; the deductive-proof test showed no signigicant change in error score after one week. In no case were significant effects, due to experimental treatments or due to an interaction between trials and treatments, found for error scores.

Properties of the Criterion Measures

One of the purposes of the present problem was to develop a satisfactory set of criterion measures of the behavior learned in the programs. As such, the measures used should be reliable and should discriminate between different levels of performance. The results of analyses of such properties are summarized here and presented in detail elsewhere (Evans, Glaser, & Homme, 1960).

Reliability. Split-half and test-retest reliability coefficients for total immediate and total retention error scores were computed. Reliabilities were computed for each experimental treatment separately, yielding a total of 12 split-half reliabilities for total test scores, one for each immediate and recall testing. Nine of the 12 split-half reliabilities coefficients were 0.92 or above, and all 12 were 0.84 or above. These findings indicate that the reliability of both the immediate and retention measures developed in this study are quite satisfactory for experimental purposes. Of the six test-retest coefficients, one for each experimental treatment, one was 0.66 while the remaining reliabilities ranged from 0.88 to 0.98. Such findings support the position that the immediate and retention tests can be treated as separate trials on the same task, thus justifying repeated-measures analysis.

Dispersion. A satisfactory measuring instrument should permit discrimination between the various objects or events to which it is applied. A performance test on which all Ss obtain perfect scores or all incorrect responses provides no range or dispersion and would obviously be unsatisfactory. An inspection of test score ranges and standard deviations, for each treatment group on both immediate and recall tests, reveals good dispersion among subjects and indicates that the purpose of measurement was adequately served by the test instruments.

Subject Characteristics and Performance

In addition to time and performance scores, several other measures of a more qualitative nature were available on each S for analysis. Analyses of the effect of sex, mathematical experience, and college class on performance are presented in this section.

Sex. In all, 27 males and 33 female Ss participated in the present study. To investigate the possible relevance of the sex variable on performance, each S's total error score was taken as the overall index of test performance. Scores were then divided into two groups on the basis of sex. The mean of the male group (35.1 errors) was tested against the mean of the female group (34.1 errors) using a one-way analysis of variance. As would be predicted from two such similar means, no significant difference in error scores attributable to sex was found.

Mathematical Experience. Records were available on the number of high-school and college mathematics courses which each S had taken. The total number of courses in mathematics which each S had taken was used as an index of "mathematical experience." These scores in turn were correlated with total error scores on all six performance tests. The overall correlation coefficient between mathematical experience and total error score was -0.11. Apparently little relationship exists between performance measures and extent of mathematical experience as indexed by number of courses in mathematics that an S had attended.

College Class. An attempt was made in recruiting to take only freshman and sophomore level Ss. However, scheduling difficulties necessitated the inclusion of several juniors, seniors and special-classes students. To check the possibility that college experience per se might be relevant to performance, a chi-square analysis of total error scores was made. Ss were divided above and below the median error score, and then cross-classified as to college class. The obtained chi-square value was not significant.

In summary, three analyses were made to determine the extent of the relationship between criterion performance on a task in symbolic logic and characteristics of Ss such as sex, mathematical experience, and college class. No such relationships were found. This appears to be evidence for the general suitability of the chosen task for experimental purposes.

The Attitude Questionnaire

A questionnaire was administered after the learning programs to check the possibility that the experimental treatments had differential effects on the attitudes of Ss toward certain aspects of this technique of learning. The results of analyses of responses made to three items on the questionnaire are presented below.

The first item sampled the attitude of Ss toward taking a course which would employ a learning program. To check the possibility that the experimental treatments themselves might affect such an attitude, a chi-square analysis was made. The "definitely like" and the "like somewhat" categories, as well as the "dislike somewhat" and "definitely dislike" categories were combined to insure adequate expected values for the cells in Table 3.10. The obtained chi-square value was not significant.

Table 3.10

Chi-square Analysis of the Relationship Between Experimental Treatment and Attitude toward Taking a Course Using Programmed Learning

Attitude	Treatment						Total
	RC	MC	IR	IF	FP	FP+	
Like	4	4	8	5	6	7	34
Indifferent	2	3	1	1	0	1	8
Dislike	4	3	1	4	4	2	18
Total	10	10	10	10	10	10	60

Chi-square = 2.99 This value is not significant for 10 degrees of freedom.

A similar chi-square analysis was performed to see if any relationship was present between experimental treatments and ratings by Ss as to whether they thought they could have learned "better, the same, or not as well" by more conventional methods of presentation. A chi-square value was not significant. In order to assess the attitude of the students toward the amount of review in the program, they were asked to judge whether the amount of review was "too much, about right, or too little." Again, the obtained chi-square value did not reach significance.

In summary, the six experimental treatments employed in the present study produced no significant effect on Ss' attitudes toward the following factors: (a) taking a course employing programmed learning; (b) effectiveness of program presentation as compared with conventional presentation of material; and (c) adequacy of amount of review in the learning programs.

Discussion

This section will be developed in the following manner. First, the implications of the findings of this study for the field of verbal learning will be discussed. Following this, some suggestions for application of these results in the area of programming technology will be presented. Finally, aspects of a standard learning program in symbolic logic for experimental investigations in verbal learning will be developed.

Implication of the Specific Findings

One very general statement can be made concerning the six experimental treatments selected for the present study. With respect to statistical significances obtained, the effect of the experimental treatments was on time scores but not on error scores. Treatment effects on time scores were present both in learning time and in time spent on criterion performance measures. Treatment effects on error scores, however, in no case approached significance, either on immediate performance scores or on the retention scores.

Keeping in mind that all treatments produced essentially the same criterion performance, consider the effect of such treatments on learning time spent on the programs. Ss making no written responses to the Initial Program (Treatment IR) finished in about 20 minutes less time, on the average, than did the two groups who composed their response to the Formal Program (FP and FP+). Mean completion times for these two groups were practically the same. The three remaining treatments (MC, IF, and RC), all requiring a written response to the

Initial Program, produced learning times from about 15 to 25 minutes longer than the two Formal Program treatments. The Initial Program contained fewer items (72) than did the Formal Program (125). However, the number of individual responses called for by the Initial Program (189) was more than the number in the Formal Program (151).

With respect to the four Initial Program treatments, it appears that by allowing implicit responding, Ss can complete their programs in about 65 percent of the time taken by Ss who must record some overt composition or multiple-choice response. This finding of less time for implicit responding is consistent with results obtained by Evans, Glaser, & Homme (1960) using a program which presented fundamentals of music.

With respect to the two Formal Program treatments (FP and FP+), availability of a review card apparently had almost no effect on mean time to complete the program. Informal observations by E during experimental sessions revealed frequent use of the review card by Ss. In addition, eight of the ten Ss under Treatment FP+ indicated that they found the card "very helpful" or "extremely helpful." The other two Ss were observed to make use of the card, but they did not record a comment on its usefulness. Since the FP+ group did spend time using the review card, and yet took about the same total time as did the FP group, the inference can be made that the review card must have increased to some degree the rate of the FP+ group on the program itself.

Additional treatment effects were in evidence for the time scores on the performance tests. Examination of these time scores reveals two facts: (a) the implicit-response group, who took from 20 to 50 minutes less learning time than the other treatments, consistently took more time on each of the three immediate performance tests; and (b) all such differences in performance times between IR and the other groups disappeared by the time of the retention tests. All treatments showed consistent drops in completion time for all tests over the retention interval, and such drops were statistically significant. The reduction in mean time for the IR groups was always greater than the corresponding reductions for the other groups. On total immediate and total retention performance time, for example, the mean time of the IR group dropped almost 20 minutes over the interval, while mean time of the other five groups were dropping two to seven minutes. Such differences in magnitude of drop appears to account for the observed trial x treatment interaction effects.

Before contrasting the results obtained on time scores with those obtained on errors scores, a brief review of the error-score findings is in order. Analysis of error scores revealed: (a) no detectable effects on criterion performance due to treatments, either on the immediate or on the retention measures; and (b) differential retention effects as a function of the type or performance measure employed. Considering time and error scores together, the following picture of the effects of the six experimental treatments in the present study emerges.

By allowing Ss to respond implicitly to program items which presented correct responses at the bottom of such items (Treatment IR), marked reductions in learning times can be obtained with no significant decrement in immediate or retention performance as indicated by error scores. Ss in this treatment apparently took time to "warm up" to overt responding on the performance tests, as reflected in longer immediate performance times for this group. In any event, such increased times to complete the performance measures were not accompanied by increased errors, and the treatment effect on time had disappeared by the end of the retention measure one week later.

Now consider the time and error scores on the two Formal Program treatments (FP and FP+). Results indicate that treatments involving overt responding to a formally-constructed program can produce, in less learning time, criterion performance comparable to that produced by treatments involving overt responses to a less systematic program. Provision of a "memory storage" device for review purposes during the learning session reduced both learning errors and criterion errors, but the reduction in the latter was not significant. The presence of a review device during learning apparently had no systematic effect on performance times.

Ss making multiple-choice responses to program items (Treatment MC) had somewhat less learning time and somewhat more performance errors that Ss who composed their answers in full to the same program.

Finally, provisions for "immediate feedback" following composed responses to the Initial Program (Treatment IF) appeared to have little effect on learning time and learning errors, as compared with a treatment which delayed such feedback from about 30 seconds to five minutes (Treatment RC). The group receiving immediate confirmation of responses made slightly more errors and took more time on the performance measures than the group whose confirmation of responses was delayed until completion of an item.

A satisfactory discussion of the obtained finding must account for the observed differences in time scores and the absence of such differences in error scores, as well as the differential retention results found for the different types of performance measures. Differences in time scores due to treatments will be presented first.

With respect to time scores, the most pronounced effects were associated with Treatment IR. As far as learning time, it is not surprising that it takes longer to compose a response, record it, and then check it than it does to compose the response and to check it without recording. Also, on more difficult items, Ss responding implicitly could prompt themselves with the correct answer immediately, while Ss responding overtly had to produce and record their response in the absence of such a prompt. On performance times, Ss who had been responding overtly continued such overt behavior, while implicit-responding Ss were writing out symbols, rules, and proofs for the first time on the criterion tests. Apparently this lack of overt practice delayed times on immediate tests. Such a practice session was adequate, apparently, to bring this rate of overt responding up to that of the other groups, as indicated in retention times.

As for the five overt-responding treatments, it will be recalled that the two Formal Program groups took less learning time, but also had fewer total responses to make than the three remaining Initial Program treatments. Without pressing the problem of the size of a verbal response unit too far, it can be stated that all five overt-response treatments averaged very close to 1.5 responses per minute, as compared with 2.3 responses per minute for the implicit-response group. In this light, the Formal Program groups appear to be responding at approximately the same overall rate as the Initial Program groups during the learning phase. Performance times for these treatments are quite similar for all six performance tests, with the possible exception of Treatment IF. On the immediate true-false and deductive-proof tests, Ss in this group took somewhat more time than Ss in the four other overt-response groups.

Two premises of modern learning theory are that organisms learn by doing and that organisms learn best when correct responses are followed by immediate confirmation or feedback (Estes, 1960). In the present study, Ss learned an overt task while responding implicitly, and delays in confirmation up to five minutes had little effect on criterion performance. The following discussion is an attempt to account for these anomalous results, and to point out what seems to to be a fundamental difference between programmed and non-programmed approaches to the study of verbal learning.

Classical techniques for the investigation of verbal learning are characterized by the precautions taken to prevent learning from occurring. Nonsense-syllable lists are standardized for low-association value; concept-formation problems contain irrelevant stimulus dimensions; problem-solving tasks are selected for their novelty. As a consequence, a characteristic of the initial stages of such learning is the number of response errors. For any increase in the probability of correct responding to occur, differential feedback as to the adequacy of such responses is obviously necessary.

In a learning program, however, the attempt is made to arrange a series of stimuli so that successive responses have a high probability of being correct from the beginning. As the learning progresses, the supporting stimuli or prompts are "faded" or withdrawn, but only at such a rate that correct responses continue to be emitted. At the termination of an "ideal" program, criterion responses should be under the control of the minimum set of stimuli which set the occasion for such responses.

Now consider the behavior of a literate adult S as he proceeds through such a program. If it is true that the stimulus portion of each item sets up a high probability that he will emit a correct verbal response, the problem of channeling such a response into any number of modalities is almost trivial. That is, if an S is adequately prepared to emit a verbal response such as "Lincoln," the correlation of responses will be almost perfect whether he is required to write it, type it, say it aloud, recognize it from a list, or write it with his toe in the sand. In the same vein, immediate confirmation of such a response should cease to be a critical factor, since S has, as it were, already confirmed the correctness of such a response himself. Evidence in support of this latter point is indicated by observations that Ss did not always turn to the back of items to ascertain the correctness of certain responses. Such items were presumably those on which Ss were confident as to the adequacy of their responses.

The preceding discussion might be generalized as follows: the relevance of variables such as response mode and immediacy of confirmation is inversely related to the probability of correct responding. That is, in situations in which correct responses have low probability, factors such as overt responding and immediate feedback are more critical than in situations in which probabilities of correct responding are high. The absence of significant effects on error scores of the four "mode-of-response" treatments (RC, MC, IR, and IF) in the present study is clearly in line with this hypothesis. Also in line are the results of an earlier study of overt versus implicit responding (Evans, Glaser, & Homme, 1960) in which no differences in performance were found following a program on fundamentals of music. Such a hypothesis would also account for the failure of Silberman and Coulson (1959) to obtain significant differences between composition and multiple-choice responding to an elementary psychology program.

For non-programmed situations such as serial, paired-associate, or multiple-choice learning in which initial correct-response probabilities are low, a prediction of the relevance of factors such as immediacy of feedback would be made. Results of a paired-associate study by Saltzman (1951), in which a six-second delay increased the number of trials to criterion by 50 percent, follow from this hypothesis. A previously mentioned study by Little (1934), in which a multiple-choice machine was used to provide knowledge of results, also confirms the necessity for immediate feedback while learning non-programmed material.

With respect to retention scores, it is interesting to note that there are three logical possibilities that a significance test of such scores can produce. Scores after a retention interval may be significantly better, they may be significantly worse, or they may not change significantly. On the true-false test, error scores decreased significantly over the retention interval. On the simple recall tests, error scores increased significantly over the retention interval. Finally, on the test calling for short deductive proofs, total scores after the retention interval were virtually the same as on the original test. It is apparent that the different types of tests showed differential effects over the one-week period. We will now consider in more detail some of the possible reasons for these effects.

As for the true-false tests, it should be pointed out that no practice was given in the programs on the specific behavior of classifying examples as being correct or incorrect instances of a logical rule. Not until the Ss reached this test itself were they required to deal with a series of possibly false examples. This proved to be a difficult task as evidenced by the large number of errors made on this test. Some Ss even scored below chance level on this 15-item exam. It is conceivable, however, that practice on this type of test is good preparation for future tests of the same type. This may account for the sizeable drop over the retention interval in the number of total errors (351 down to 232). The poor immediate performance on the true-false test may imply that to produce effective behavior on exams of the true-false and multiple-choice types, the program itself must provide specific practice on such items.

The second test in the series required Ss to recall a rule, given its name, and apply it to one or more given steps. On the retention tests the total number or errors increased from 247 to 287, or an average of about 0.67 errors per S. However, the high correlation between immediate and retention scores resulted in this increase in errors reaching statistical significance, since the experimental design permitted removal of variability due to Ss.

On the deductive-proof test, the increase in total errors after the retention interval was negligible (477 to 481). This increase failed of course to reach statistical significance. It might be noted that the deductive-proof test also involved recall and application of the logical rules, as did the recall test itself. In many instances, however, Ss were able to apply a certain rule correctly in constructing a proof even though they had been unable to recall and apply the same rule when its name was given. Apparently Ss sometimes remembered the operations involved in a rule but had difficulty in recalling the name for that operation. This would account for the slight but significant increase in errors when the operations had to be recalled, given only the name, as in the recall test. The fact that ability to perform the operations involved in a rule suffered no particular loss over the retention interval is reflected in the similarity between immediate and retention performance on the deductive-proof test.

Considering the retention data together, it appears that there is no marked decrement in performance over the retention interval, even though in the recall test an average increase of less than one error per S proved to be statistically significant. The decrease in true-false error scores over the interval is perhaps attributable to the practice effect received in taking the immediate true-false test, a type of performance not practiced in the program itself. Absence of any pronounced increase in the number of retention errors indicates that the behavior produced by the learning programs was present in approximately the same strength after the one-week interval. If this was the case, then the facilitating effect of having taken three very similar immediate tests should reduce the completion times on the retention tests. The finding of consistent and significant reductions in completion times over the retention interval is evidence for this conclusion.

In summary, the chief implication of the results of the present study for the area of verbal learning is as follows. Failure to obtain performance decrements attributable to variables such as non-overt responding and delay of feedback necessitated a re-examination of the nature of such variables in programmed learning. A distinction was made between situations in which probabilities of correct response were high throughout the learning period and situations with low initial probabilities of correct responding (e.g., nonsense-syllable lists, concept formation). It was hypothesized that the relevance of variables such as response mode and immediacy of confirmation was inversely related to the probability of correct responding. Results of both programmed and non-programmed verbal learning studies which support this conclusion were pointed out.

Program Technology

Some implications of the present investigation for the area of "program technology" will now be presented.-- The main result of this study of import for programming methodology was the demonstration that specifiable sequences of "standard" item-types (e.g., rule-example-incomplete-example) can produce, in less learning time, the same level of criterion performance as programs constructed according to less formal principles and procedures. Such results, if confirmed by other studies in other topics, could lead, in the long run, to procedures for programming knowledge for humans which are as rigorous and systematic as programming procedures for digital computers.

More immediately, the ease with which items can be added to the Formal Program could be used to expand it to improve criterion performance. Of the programming variables reported in the literature, the "size-of-step" variable has been most consistently related to improving performance (Evans, Glaser, & Homme, 1960; Silberman & Coulson, 1959). Since the two Formal Program groups (FP and FP+) had learning times from 20 to 30 minutes less than the three overt-responding Initial Program groups (MC, IF, and RC), such time could be used to present additional items.

Another procedure for increasing the number of items in a program is associated with the finding of significantly less time for the implicit-response treatment (IR). Again, since savings in time can be accomplished without performance decrement by allowing implicit responses, programs could be expanded so that total learning time is approximately the same as for overt-response treatments. For the present task, a combination of an expanded Formal Program with implicit responding should permit learning time to be reduced by one-half as compared with overt responding to the Initial Program. Under these circumstances, a large number of additional items or steps could be added without increasing the average learning time over the somewhat arbitrary two-hour time limit.

A disadvantage of allowing implicit responding for programming research is that S leaves no record of his responses. As Skinner (1958) has pointed out, a salient feature of learning programs is the progressive modification and improvement following analyses of recorded responses. In the developmental phase of a program, the necessity for recording S's responses still remains. After a program is producing satisfactory criterion behavior, it may be desirable to require recorded responses to some items but to permit implicit responding to the remaining items. If such a procedure produced no performance decrements, appreciable savings in completion times may be gained without loss of assurance that S is responding correctly. In addition, the reduced number of recorded responses would consume less space and perhaps permit simplification of the device which receives such recorded responses.

Another result that has possible implications for program technology comes from a comparison of response-composition (RC) and multiple-choice (MC) responding to the Initial Program. Total performance errors under Treatment MC were 25 percent greater than total performance errors under Treatment RC. Because of large within-group variability such differences are not statistically significant, but a mean difference of 7.5 errors per S between the treatments deserves some comment. Such a difference favoring composed or constructed responses would be predicted by Skinner (1958), who states that incorrect multiple-choice answers on an item may compete with the correct response. However, if the hypothesis relating correct-response probabilities and mode of response holds, it would follow that as the probabilities of correct responding on a program increase as the program is successively improved, response-mode differences should decline. For the present program, low correct-response probabilities were present on later items, as evidenced by incorrect responses made by many Ss. In this case it is possible that response-competition and interference effects on MC items did occur, with consequent performance errors. As the present program is expanded and improved, the difference in mean number of errors between RC and MC treatments should decline.

A brief comment is also in order as to the effect of the "memory storage" device used in Treatment FP+. This treatment, involving the Formal Program plus the use of a review card containing the logical rules, produced the fewest total performance errors of the six treatments studied. It is interesting to note that Ss on the same program without the card (FP) made 15 percent more total performance errors, and about 50 percent more learning errors than the FP+ group. Again, within-group variability prevented the differences in performance between these two groups from being statistically significant. However, the fact that provision of a summary type prompting device reduced both learning and performance errors deserves further study.

With respect to the implications of the present investigation, several interesting possibilities arise. Since delay of confirmation from 30 seconds to five minutes resulted in no significant performance decrement, what would happen if such confirmation were withheld altogether? If the previous analysis of situations involving high correct-response probabilities is correct, it is possible that satisfactory performance can be attained without providing confirmation at all. A possible complication here is that such external confirmation may not be necessary for learning per se but may be necessary to motivate such behavior over the course of a program. That is, the primary function of the confirming stimulus may not be the strengthening of the response just emitted. Rather, such stimuli may serve to maintain such responding until the program is completed.

Another related variation involves the procedure used in the implicit-response treatment(IR). It will be recalled that the correct answers to each item were removed from the context of the item and placed at the bottom of that item. An alternative procedure would be to leave such answers and solutions in context. Ss in such a treatment would presumably make their own implicit responses to items in much the same manner as they would when studying from a text. Results on performance measures would indicate whether or not the "blanks" and spaces which signal responses in a program are necessary to direct the attention of Ss to the critical aspects of that item.

Another question can be raised with respect to implicit responding. Much verbal behavior is so complex that Ss appear to need the stimulus support provided by their own recorded responses to complete a response correctly. Examples of this would be drawing a complicated electrical circuit or sketching a complex organic compound. When "units" of behavior of this size constitute the criterion behavior, overt responding during the learning phases may be necessary. The ease with which complex problems can be generated for the logic task used in the present investigation provides a technique by which the relation between response complexity and implicit responding could be studied.

A "Standard" Learning Program

With respect to the choice of symbolic logic as a topic, the advantages of such a task for problem-solving studies outlined by Moore and Anderson (1954) held equally well when the task was presented using a learning program. For example, no assumptions of previous training were made. Of all Ss who participated in the present study, only one S indicated that he found the task too difficult and asked to terminate the learning session. As has been pointed out, no relationships were observed between performance on the task and individual variables such as mathematical experience, sex, and college class. A detailed discussion of various aspects of the symbolic logic program as a "standard" learning program and recommendations for its uses in research are given in Evans, Glaser, & Homme (1960).

Summary and Conclusions

Six independent groups of ten college students each received learning programs of the "teaching machine" type. The programs were designed to teach the construction of short deductive proofs involving 15 rules in symbolic logic. Two experimental treatments involved a systematic program in which both the type and sequence of items followed the same pattern for each of the rules. Both groups using this program composed or constructed their answers to each item. One group did, and one group did not, use a review card containing all the logical rules. The remaining four treatments used a less systematic program previously developed. Four different modes of responding to the items of the program were used. One group wrote out each of their responses to items in the program. A second group also composed their answers, but received immediate knowledge of results on items involving more than one response. A third group had the correct answer present on the front of the item and were not required to make an overt written response. A fourth group selected the correct response from a set of multiple-choice answers at the bottom of the item.

A true-false test, a test involving recall of each of the rules, and a test requiring short deductive proofs were constructed to sample different aspects of the behavior learned. These tests were administered after the experimental learning sequence, and three parallel retention tests were given after a period of one week. An attitude questionnaire toward the experiment was also administered after the learning session.

Dependent measures were: time spent on the learning programs, time spent on the performance tests, and number of errors made on the performance tests.

The following conclusions are drawn on the basis of analysis of the data obtained.

1. Experimental variations in mode of responding significantly affect learning time. Ss not required to make an overt written response to each item can complete a learning program in about 65 percent of the time required for composed or multiple-choice responding.

2. Criterion performance in terms of error scores is not significantly affected by mode of responding, including no overt responding at all.

3. Systematically constructed programs can produce, in less learning time, criterion performance comparable with that of a less systematic program.

4. Ss who respond non-overtly to learning programs take significantly more time on performance tests which immediately follow the program than do Ss who make their responses overtly. Such differences in test-time disappear after a retention period of one week.

5. Differential retention effects were observed as a function of the type of criterion performance measured. Error scores on true-false tests decreased significantly; error scores on recall tests showed slight but significant increases; on tests involving deductive proofs no significant changes were observed.

6. No significant relationships are observed between performance following the programmed learning sequence employed and sex, mathematical experience, or college class.

7. Implications of the results for the area of verbal learning were discussed. It was hypothesized that the relevance of variables such as response mode and immediacy of feedback are inversely related to the probability of correct responding.

8. Suggestions for the use of programmed techniques for the investigation of problem-solving and concept-formation behavior were presented.

9. Development of a standard learning program for experimentation in the area of programmed learning was described.

References

Ambrose, A. & Lazerowitz, M. Fundamentals of symbolic logic. New York: Rinehart, 1948.

Anderson, S. B. Analysis of responses in a task drawn from the calculus of propositions. NBL Memorandum Report 608, 1956.

Bernays, P. Axiomatische Untersuchung des Aussagen-Kalkuls der Principia Mathematica. Mathematische Zeitschrift, 1926, 25, 305-320.

Briggs, L. J. Intensive classes for superior students. J. educ. Psychol., 1947, 38, 207-215.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. A study of thinking. New York: Wiley, 1956.

Carr, W. C. Self-instructional devices: A review of current concepts. Wright Air Development Center, 1959. TR-59-503.

Copi, I. M. Symbolic logic. New York: Macmillan, 1954.

Culbertson, J. T. Mathematics and logic for digital devices. Princeton: Van Nostrand, 1958.

Edwards, A. L. Experimental design in psychological research. New York: Rinehart, 1956.

Estes, W. K. Learning. In C. W. Harris (Ed.), Enclopaedia of educational research. New York: Macmillan, 1960. Pp. 752-770.

Evans, J. L., Glaser, R., & Homme, L. E. A preliminary investigation of variations in the properties of verbal learning sequences of the "teaching machine" type. In A. A. Lumsdaine & R. Glaser (Eds.), Teaching machines and programmed learning: A source book. Washington, D. C.: National Education Association, 1960. Pp. 446-451.

Evans, J. L., Glaser, R., & Homme, L. E. The Ruleg System for the construction of programmed verbal learning sequences. A report to the Office of Education, University of Pittsburgh, April 1960.

Eves, H. & Newson, C. V. An introduction to the foundations and fundamental concepts of mathematics. New York: Rinehart, 1958.

Ferster, C. B. & Sapon, S. M. An application of recent developments in psychology to the teaching of German. Harvard educ. Rev., 1958, 28, 58-69.

Galanter, E. The ideal teacher. In E. Galanter (Ed.), Automatic teaching: The state of the art. New York: Wiley, 1959. Pp. 1-11.

Gilbert, T. F. An early approximation to principles of continuous discourse, self-instructional materials. A report to Bell Telephone Laboratories, Inc., Murray Hill, New Jersey, 1958.

Glaser, R. Christmas past, present, and future. Contemp. Psychol., 1960, 5, 24-28.

Glaser, R., Homme, L. E., & Evans, J. L. An evaluation of textbooks in terms of learning principles. In A. A. Lumsdaine & R. Glaser (Eds.), Teaching machines and programmed learning: A source book. Washington, D.C.: National Education Association, 1960. Pp. 437-445.

Homme, L. E. & Glaser, R. Relationships between the programmed textbook and teaching machines. In E. Galanter (Ed.), Automatic teaching: The state of the art. New York: Wiley, 1959. Pp. 103-107.

Homme, L. E. & Glaser, R. Problems in programming verbal learning sequences. In A. A. Lumsdaine and R. Glaser (Eds.), Teaching machines and programmed learning: A source book. Washington, D.C.: National Education Association, 1960. Pp. 486-496.

Jensen, B. T. An independent study laboratory using a self-scoring test. J. educ. Res., 1949, 43, 134-147.

John, E. R. & Miller, J. G. The acquisition and application of information in the problem-solving process: An electronically operated logical test. Behav. Sci., 1957, 2, 291-300.

Jones, R. S. Integration of instructional with self-scoring measuring procedures. Abstr. doct. Dissert., 1954, 65, 157-165.

Little, J. K. Results of use of machines for testing and for drill, upon learning educational psychology. J. exp. Educ., 1934, 3, 45-49.

Luh, C. W. The conditions of retention. Psychol. Monogr., 1922, 31, No. 142.

Lumsdaine, A. A. Teaching machines and self-instructional materials. Audio-vis. Comm. Rev., 1959, 7, 163-181.

Lumsdaine, A. A. & Glaser, R. (Eds.) Teaching machines and programmed learning: A source book. Washington, D.C.: National Education Association, 1960.

Melton, A. W. Some comments on "The impact of advancing technology on methods of education," by Dr. Simon Ramo. Paper read at Amer. Psychol. Ass., Cincinnati, September 1959.

Moore, O. K. & Anderson, S. B. Modern logic and tasks for experiments in problem-solving behavior. J. Psychol., 1954, 38, 151-160.

Newell, A., Shaw, J. C., & Simon, H. A. Elements of a theory of human problem solving. Psych. Rev., 1958, 65, 151-166.

Nicod, J. A reduction in the number of the primitive propositions of logic. Proc. Cambridge Phil. Soc., 1916, 19, 32-42.

Porter, D. A. A critical review of a portion of the literature on teaching devices. Harv. educ. Rev., 1957, 27, 126-147.

Porter, D. A. Teaching machines. Harv. grad. Sch. of educ. Ass., 1958, 3, 1-5.

Pressey, S. L. A simple apparatus that gives tests and scores - and teaches. Sch. & Soc., 1926, 23, 373-376.

Pressey, S. L. A machine for automatic teaching of drill material. Sch. & Soc., 1927, 25, 549-552.

Pressey, S. L. Development and appraisal of devices providing immediate scoring of objective tests and concomitant self-instruction. J. Psychol., 1950, 29, 417-447.

Reichenbach, H. Elements of symbolic logic. New York: Macmillan, 1947.

Rosser, J. B. Logic for mathematicians. New York: McGraw-Hill, 1953.

Saltzman, C. J. Delay of reward and human verbal learning. J. exp. Psychol., 1951, 41, 437-439.

Silberman, H. F. & Coulson, J. A draft summary of findings in an exploratory teaching machine study. Automated Teaching Bull., 1959, 1 (2), 35-37.

Simon, H. A. & Newell, A. The simulation of human thought. Paper given at Current Trends Conference, University of Pittsburgh, 1959.

Skinner, B. F. The behavior of organisms: An experimental analysis. New York: Appleton-Century-Crofts, 1938.

Skinner, B. F. The science of learning and the art of teaching. Harv. educ. Rev., 1954, 24, 86-97.

Skinner, B. F. Verbal behavior. New York: Appleton-Century-Crofts, 1957.

Skinner, B. F. Teaching machines. *Science*, 1958, *128*, 969-977.

Skinner, B. F. *Cumulative record*. New York: Appleton-Century-Crofts, 1959.

Smith, D. E. P. Speculations: Characteristics of successful programs and programmers. In E. Galanter (Ed.), *Automatic teaching: The state of the art*. New York: Wiley, 1959. Pp. 91-102.

Whitehead, A. N. & Russell, B. *Principia mathematica*. New York: Cambridge University Press, 1925.

PRECEDING PAGE BLANK-NOT FILMED

Chapter 4

A Study of the Effects of Partial Reinforcement on Behavior in a Programmed Learning Situation[1]

A major principle of learning concerns the effects of the consequences of behavior on the behavior itself; when an organism is reinforced, in some way, for engaging in a particular behavior, the response made, depending upon previous reinforcement history, will undergo a change in strength. Of particular interest is the effect of supplying reinforcement on a non-continuous schedule. When a response is reinforced on an other than continuous schedule, the response will tend to attain a higher strength than if it were reinforced upon every occurrence. The scheme by which a reinforcer is "dispensed," in terms of the number of reinforcers given to a number of responses, is called a schedule of reinforcement. Any schedule which calls for fewer reinforcements than responses (a ratio of less than 1:1) is called a partial schedule of reinforcement. Ferster and Skinner (1957) list no less than 14 different schedules of reinforcement upon which research has been done. The Ferster and Skinner work is a technical manual which describes research on schedules of reinforcement where a free operant was employed. A review of the partial reinforcement literature concerned with discrete trial behavior has recently been presented by Lewis (1960). An earlier review (Jenkins and Stanley, 1950) covering work prior to 1950 and including some human subject studies is also available. Most of the general comments below are based on these references.

Research interest in schedules of reinforcement has taken three major forms. The first of these is represented by the work of Ferster and Skinner which deals with the behavioral consequences of imposing various schedules in terms of frequency and temporal patterning of behavior. A second, and closely allied interest, is the study of parametric situations in which some independent variable is related to response acquisition and resistance to extinction. This latter interest is exemplified by Sheffield's (1949) study in which rats were studied under conditions of massed and spaced learning, and different groups of animals were presented with different percentages of reinforcement. A third type of study in partial reinforcement is the investigation of partial reinforcement effects in an effort to substantiate or refute some point of theory. Typical of the partial reinforcement literature is the use of animal subjects (rat, pigeon, or monkey) and the use of precisely defined response classes such as lever pressing. Few studies have used human subjects and none have attempted to apply a partial reinforcement situation to the type of learning done in the classroom. Classic studies of partial reinforcement at the human level are those of Humphreys (1939; 1940) in which he compared 100 percent reinforcement with 50 percent reinforcement for conditioned eyelid and psychogalvanic responses and found that conditioning was equally rapid in both cases and that resistance to extinction was greater for the 50 percent groups. With regard to extinction, of the 17 separate studies dealing with partial reinforcement in humans cited by Jenkins and Stanley, only three failed to find the partially reinforced subjects displaying greater resistance to extinction.

---

[1] The data presented in this report was originally submitted by Eugenia Stillman Scharf to the Graduate School, Division of Natural Sciences, University of Pittsburgh (July 1961) in partial fulfillment of the requirements for the degree of Master of Science.

In attempting to translate a basic research finding such as that of the effects of partial reinforcement into an applied study, a number of serious problems are immediately encountered. The desire, of course, is to be able to apply research findings to the world of practical affairs. It is difficult to select a class of behavior for study which will provide a relatively pure measure of the effects of experimental manipulations while at the same time providing a basis for generalization to applied problems. As a task in the present study, the symbolic logic program previously developed and studied by Evans, Glaser, and Homme (see Chapter 3) was selected. This task provides measurable behavior that is relatively easy to specify and bears formal similarity to the kinds of problem-solving learning of interest in education. A further consideration involves the selection of a measure of response strength. Traditionally, either rate of learning or extinction has been used as a strength measure in animal studies of partial reinforcement. The closest the present study could come to an extinction session was the use of a 30-item test (unreinforced responding) following learning.

As pointed out in Chapter 3, most standard learning tasks for laboratory use involve considerable incorrect behavior early in learning. This is in contrast to the intention of programmed learning where the learning situation is arranged to minimize the occurrence of erroneous behavior. In differentiating between these two kinds of learning situations, Chapter 3 suggests the hypothesis that the high probability of a response being correct (as should be the case in a programmed learning sequence) should tend to decrease the relative effects of different percentages of reinforcement during acquisition. When a response is highly probable, the effect of and need for reinforcement is less important than when the response is at low strength. Hence, the influence of variations in reinforcement conditions should be less apparent in high probability programmed learning situations than in low probability situations such as studied in the literature reported.

The purpose of the present study was to observe the effects of varied percentages of reinforcement on groups of subjects using a standardized programmed learning task. Reinforcement is here defined as the availability of the correct response following the emission of a response to a program frame.

Method

Apparatus. In selecting an appropriate program to use as the apparatus in this study, several requirements had to be satisfied. The program had to be brief, it had to be a finite piece of behavior, and it had to be simple enough for an unselected high school group. The program was to be used primarily as an instrument to study the effects of experimental variables, and the end result of the program had, therefore, to be a measurable behavior. High school juniors were selected as subjects for the study because of their ready availability.

A number of changes in Evans' symbolic logic program were made to enhance its value as a learning task. The program was written as a sort of game. It consisted of three sections and a 30 item post-test. Using a variation of the programmed textbook developed by Homme and Glaser (1959), the program was printed on 5" x 8" cards, one frame per card, with the correct answers on the reverse side. The program was constructed by means of the Ruleg System described in Chapter 2.

Section I of the program contained 55 cards designed to familiarize the subjects with the requisite terminology and methods of approach. The frames dealt with the operators and elements used in performing symbolic logic operations. Section II consisted of 81 cards dealing with ten of the rules of operation; that is, it defined the type of manipulations permissable. In Section III, the subjects were taught "How to Get to W," W being the Winning Step or the end result of the manipulations. At the end of this section, the subjects were performing short deductive proofs. These two-step proofs constitute the criterion behavior, and all 30 items on the post-test were various examples of this behavior.

As each section of the program was being developed, it was tried out and revised on the basis of subject performance. In an effort to shorten the required time, and on the basis of previous findings (Chapter 3), implicit responding was permitted to approximately 30 percent of the items in the first section. As in the previous study, no differences were found between subjects writing down every response and subjects with only partial overt responding. The average error in the entire first section was found to be five percent. A criterion test consisting of 29 items on five cards had been built into this section. The average error on this test alone was two percent. The average error in Section II in the preliminary study was eight percent and in Section III it was two percent. The average number of errors made on the 30 item post-test was 4.7 or 16 percent. The reliability of the post-test, using all subjects described below was found to be .90, using the Kuder-Richardson formula 21.

The length of the entire program prevents its inclusion in the present report and important details of its construction have been reported elsewhere (Scharf, 1961). The program, as used in this study, was quite similar to the one previously developed by Evans, Glaser, and Homme (Chapter 3) and adapted by Hessert (Chapter 5). Whereas the earlier program taught 15 rules of logic to college subjects, the present program taught ten such rules in a somewhat simplified manner to high school subjects. As used in this study, the entire program contained 166 frames divided into three sections.

College freshmen and high school students from a well-above average school were used in building the present version of the program. Thus, the program was standardized only for a high-level high school group. As could be expected, the program did not work well with the lower I.Q. students and, thus, the data shall be discussed mainly in terms of the upper half of the experimental subjects.

Subjects. Eighty-four high school juniors were selected as subjects. All were taking an academic course in a middle income high school. Absences during completion of the experiment reduced the total number of subjects to 69, 27 males and 42 females. Prior to the selection of subjects, the most recent Otis I.Q. scores of all members of the junior class participating in the academic program were obtained and listed in order from high to low. Subjects were selected by first taking every fourth name on the list and assigning every other selected name to one of two groups. These groups were then further sub-divided so that there were four matched groups for use in the experiment. The original 84 subjects were thus divided into four groups of 21 subjects each. The mean and range of the four groups are shown in Table 4.1.

Table 4.1

Final Means and Range of Otis I.Q. Scores for the Four Reinforcement Groups

	Group			
	I	II	III	IV
Mean I.Q.	110	108.2	107.7	111.6
Range	89-136	94-124	85-134	92-127

Procedure. All four groups went through Section I of the logic program in the same way. All responses were reinforced. The subjects, however, wrote out their responses only to certain pre-specified items. The subjects were required to mark with an "X" all written incorrect responses made on this section. In an attempt to control cheating (i.e., looking at the correct answers on the back of the card before writing out the response) subjects were instructed to go through the cards in sequence and not to return to any preceding cards. The information presented in the first section is basic to any mastery of the subject matter. Therefore, all groups were given the same background prior to the introduction of reinforcement variations.

The experimental variations took place in Sections II and III. All subjects in all four groups responded overtly to all items. This was a practical requirement, as almost all of the responses in these sections involved mathematical symbols and manipulations, as well as some fairly long verbal responses. It was no longer necessary for S to mark his incorrect answers.

Group I received 100 percent continuous reinforcement; that is, all answers were given on the back of the cards. Group II received 50 percent fixed ratio reinforcement (FR_2); that is, the answer to every other item was given. Group III received 50 percent variable ratio reinforcement (VR_2); i.e., the answers to one-half of the items were given but in no particular patters. Group IV received 25 percent variable ratio reinforcement (VR_4). It is apparent that these are arbitrarily selected schedules of reinforcement, and that the most effective schedule might well fall somewhere in between. It was reasoned, however, that if no significant differences were found between these arbitrary extremes, it would be highly unlikely that they would be found with other ratios.

The same post-test was given to all groups. The program was administered in three sessions: Section I on a Monday, Section II on a Wednesday, and Section III and the post-test on Friday. Each session was approximately two hours long including organization and rest periods.

Treatment of the Data and Results

The data will be discussed in terms of four measurements: (a) errors made on the criterion test, (b) errors made on the program, (c) time taken to complete the criterion test, and (d) time taken to go through the program.

Each of the four reinforcement groups was divided into a high and a low I.Q. subgroup. This was done for two reasons: (a) the program was not standardized on a low I.Q. group and, therefore, might not work well for such a group. Thus, by not separating the subjects in this way, pertinent data might be confounded; and (b) it was of interest to note the magnitude of the differences between these two groups on the above measures. The data obtained, in terms of means for the various experimental groups, are shown in Table 4.2.

Table 4.2

The Experimental Design

		Amount of			Sections I-III		Post-test	
	Group	Reinforcement	Mean I.Q.	N	Errors	Time	Errors	Time
High	I	100%	120	7	19.643	188.428	5.714	49.857
I.Q.	II	50% fixed	115	7	24.428	208.714	9.143	49.428
Group	III	50% variable	117	8	28.312	192.250	9.500	50.375
	IV	25% variable	117	11	25.909	179.909	9.272	48.545
Low	I	100%	102	9	40.222	194.444	14.000	44.444
I.Q.	II	50% fixed	102	8	49.562	204.500	17.750	44.375
Group	III	50% variable	101	11	56.454	192.454	17.682	45.454
	IV	25% variable	104	8	63.562	188.875	18.188	45.000

After applying Bartlett's homogeneity of variance test, and finding variances to be homogeneous, four major analyses of variance were performed. Because of the desproportionate n's, the method of unweighted means was used (Bendig, 1960). The mean of all the subjects was determined and found to be 109.43. All Ss with I.Q.'s above this mean constituted the high I.Q. group; all Ss with I.Q.'s of 109 or less constituted the low I.Q. group. Two analyses per measure were needed: the first to determine the gross between and within group variance; the second to determine whether the between group variance was attributable to the effects of the differential reinforcement or to the different I.Q. subgroups or, of course, to both.

Table 4.3 indicates that the low I.Q. subjects in each of the four reinforcement groups made approximately twice as many errors on the post-test as did the high I.Q. subjects.

Table 4.3

Mean Number of Errors (30 items) on Post-test

			Reinforcement Group	
	100%	50% fixed	50% variable	25% variable
High I.Q.	5.714	9.143	9.500	9.272
Low I.Q.	14.000	17.750	17.682	18.188

From Table 4.4 it may be seen that there was no significant relationship between schedules of reinforcement and number of errors made on the post-test. Group I, however, the 100 percent reinforcement group, made far fewer errors than did those groups receiving partial reinforcement. The implications of these findings will be discussed in the following section. As expected, the differences between the scores made by the high and low I.Q. subgroups, regardless of reinforcement group, was highly significant.

Table 4.4

Analysis of Variance of Post-test Error Scores

Source of Variation	df	Uncorrected Sum of Squares	Corrected Sum of Squares	Mean Square	F
Total	68	4233.79			
Between Groups	7	1371.71		195.96	4.18**
I.Q.	1	144.24	1211.610	1211.610	25.82****
Groups	3	20.973	175.948	58.649	1.25
Interaction	3	.165	1.384	.461	
Within Groups	61	2862.08	2862.08	46.919	

**p < .01

****p < .0001

Tables 4.5 and 4.6 indicate that, as on the post-test, the low I.Q. groups made approximately twice as many errors on the program itself as the high I.Q. groups regardless of reinforcement group. Again, there was no significant relationship between schedule of reinforcement and the number of errors made on the program itself. The 100 percent reward group, however, again tended to make fewer errors than the other groups. The difference between the high and low I.Q. subgroups was significant.

Table 4.5

Mean Number of Errors (203 items) on Sections I - III of the Program

	Reinforcement Group				
	100%	50% fixed	50% variable	25% variable	Sum
High I.Q.	19.643	24.428	28.312	25.909	98.292
Low I.Q.	40.222	49.562	56.454	63.562	209.800
Sum	56.865	73.990	84.766	89.471	308.092

Table 4.6

Analysis of Variance of Program Errors in Sections I - III

Source of Variation	df	Uncorrected Sum of Squares	Corrected Sum of Squares	Mean Square	F
Total	68	55,103.81			
Between Groups	7	15,987.62		2,283.94	3.56**
I.Q.	1	1,554.331	13,039.687	13,039.687	20.33****
Groups	3	259.329	2,175.578	725.193	1.13
Interaction	3	78.134	655.486	218.495	
Within Groups	61	39,116.19	39,117.19	641.25	

** p < .01

**** p < .0001

Tables 4.7 and 4.8 show that there was no significant relationship between reinforcement groups and the amount of time taken to complete the post-test. There was, however, a significant difference between the low and the high I.Q. subgroups, the former taking significantly less time to finish the test.

Table 4.7

Mean Time (Minutes) to Complete Post-test

Reinforcement Group

	100%	50% fixed	50% variable	25% variable	Sum
High I.Q.	49.857	49.428	50.375	48.545	198.205
Low I.Q.	44.444	44.375	45.454	45.000	179.273
Sum	94.301	93.803	95.829	93.545	377.478

Table 4.8

Analysis of Variance of Mean Times Spent on Post-test

Source of Variation	df	Uncorrected Sum of Squares	Corrected Sum of Squares	Mean Square	F
Total	68	4,637.77			
Between Groups	7	386.77		55.25	
I.Q.	1	44.800	375.839	375.839	5.393*
Groups	3	1.574	13.205	4.402	-
Interaction	3	1.007	8.448	2.816	
Within Groups	61	4,251.000	4,251.000	69.688	

*$p < .05$

In regard to time taken to complete the program (Sections I - III), Tables 4.9 and 4.10 show two unusual findings: (a) I.Q. had no effect at all, and (b) there was a significant difference between the reinforcement groups, Group II (50 percent variable reinforcement) taking significantly longer to go through the program than did the other three groups. To determine how significantly this group differed from the others and how the others differed among themselves, six t-tests were run. From Table 4.11, we can see that the mean of Group II (the 50 percent variable reinforcement group) differed significantly from the means of all of the three other groups.

Table 4.9

Mean Time (Minutes) Spent on Sections I - III of the Program

	Reinforcement Group				
	100%	50% fixed	50% variable	25% variable	Sum
High I.Q.	188.428	208.714	192.250	179.909	769.301
Low I.Q.	194.444	204.500	192.454	188.875	780.273
Sum	382.872	413.214	384.704	368.784	1549.574

Table 4.10

Analysis of Variance of Mean Time Spent on Sections I - III on the Program

Source of Variation	df	Uncorrected Sum of Squares	Corrected Sum of Squares	Mean Square	F
Total	68	21,386			
Between Groups	7	4,980		711.43	2.64*
I.Q.	1	15.048	126.242	126.242	
Groups	3	520.346	4,365.318	1,455.106	5.410**
Interaction	3	52.140	437.416	145.805	
Within Group	61	16,406	16,406	268.95	

* $p < .05$
** $p < .01$

Table 4.11

Groups

Groups	Mean	100%	50% fixed	50% variable	50% variable
100%	191.436		2.574**	.165	1.266
50% fixed	206.607			2.517**	3.922***
50% variable	192.352				1.496
50% variable	184.392				

** $p < .02$
*** $p < .01$

Discussion and Implications

The differences in mean error scores made by the four reinforcement groups on both the program itself and on the post-test were not statistically significant. This departure from the usual finding that 100 percent reinforcement is more effective than partial reinforcement during acquisition can be explained in terms of the nature of programmed learning. As previously mentioned, in programmed learning each frame has a high probability of eliciting the correct response. The effect of continuous reinforcement, therefore, cannot be assumed to be the same as it is when less familiar stimuli which can elicit a large number of responses, are used.

In accordance with previous findings, however, there was a definite trend showing the 100 percent reinforcement group to make fewer errors than any of the partially reinforced groups. The reasons for this are somewhat unclear. The trend might be due to the obvious and often demonstrated fact that 100 percent reinforcement is more effective than partial reinforcement in the early stages of learning, or it might be due to the increased exposure to the material provided by the repetition of items constituting the reinforcement. In the latter case, however, there should then have been an equivalent discrepancy in error scores between Groups III and IV as for Groups I and II. However, the latter group received only one-half as much repetition of the material as did the former (i.e., 25 percent as compared to 50 percent) and no such differences were found.

There appeared to be differences between the four groups in time taken to complete the post-test. There was, however, a significant difference in time taken for the program itself, Group II taking significantly longer than any of the other groups. Why this is so is difficult to explain. In terms of the additional reading matter (i.e., the confirmation of response), this group had half as much as did Group I and no more than did Group III. This group took approximately the same length of time to complete the post-test as the others.

It is relevant to examine the implications of this study with respect to the nature of reinforcement in programmed learning. The fact that confirmation of response is an adequate reinforcement has been generally accepted. Apparently, 100 percent reinforcement is helpful but not completely necessary. It would appear that overt confirmation of answers in such a high probability task is not particularly essential at all. It might be hypothesized that the major, over-all reinforcement in such learning is simply the completion of the task itself and, as

Skinner (1954) has suggested, that "exploratory and manipulative activities may be reinforcing enough to keep a student at work." In accord with such a hypothesis, Holland (1960), using a programmed course in psychology, randomly divided Harvard students into three groups. Group I used the regular program, i.e., they wrote down their responses, operated the machine and were duly reinforced before going on to the next frame. Group II also wrote down their responses but did not receive any confirmation. The third group simply read through the completed statements. All blanks were previously filled in, and they neither responded nor were they reinforced. In a test immediately following, Group I had a mean error of seven percent, Group II of nine percent, and Group III of 15 percent. The slight difference between the first two groups appears to illustrate that in such high probability learning, confirmation of response is not necessary. "The student may not need to be told he is right, since it is usually obvious to him," (Holland, 1960).

As expected, particularly on the basis of the standardization of the logic program, it was found that the effect of I.Q. level made important differences in this study. I.Q. was significant in every measure but one (time taken to complete the program). The difference between high and low I.Q. groups on number of errors made was significant regardless of reinforcement group. In general, the brighter Ss were easily able to grasp the meaning, as well as the rudiments of the task. They made some errors, of course, but were usually able to correct their wrong responses in a short time. That is, they profited from the "non-confirmation" of their wrong answers. In contrast, it is obvious that this program was definitely not suited for the majority of the low I.Q. group. Some apparently ignored the examples as well as the correct answers on the back of the cards. Thus, the same wrong responses were made over and over again. The errors made by this group were, in general, much more pervasive and of longer duration than those made by the high I.Q. group.

These results produced by I.Q. differences suggest a rather basic problem of program construction. Skinner, for example, apparently felt that bright and dull students should progress equally well and should be treated the same in terms of teaching machine programs. Pressey, on the other hand, suggested separate programs for each of these groups. It is possible, of course, through extensive repetition and review to assure that even the dullest subject will get through a program with a minimum of error. But such an achievement might occur at the expense of learning efficiency for the brighter students even though they would go through the program much faster than the slower students.

In conclusion, it would seem that continuous reinforcement may notbe a *sine qua non* of a good program. Other studies, using other types of programs and subjects should be carried out. It is possible that partial reinforcement works as well as continuous reward only with older subjects and with specific subject materials. Younger children, on the other hand, might benefit more from continuous reinforcement. In view of its advantages of partial reinforcement as a technical time-saver and in view of its past proven superiority to continuous reward in terms of resistence to extinction (a variable which should also be studied in programmed learning), partial reinforcement techniques might have an important part in the future development of programmed learning.

References

Bendig, A. W. Analysis of variance with unequal subclass frequencies. University of Pittsburgh, March 1960. (Mimeo)

Ferster, C. B. & Skinner, B. F. *Schedules of reinforcement*. New York: Appleton-Century-Crofts, 1957.

Holland, J. G. Design and use of a teaching machine program. Paper presented at Amer. Psychol. Ass., September, 1960.

Homme, L. E. & Glaser, R. Relationships between the programmed textbook and teaching machines. In E. Galanter (Ed.), *Automatic teaching: The state of the art*. New York: Wiley, 1959. Pp. 103-107.

Humphreys, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, *25*, 141-158.

Humphreys, L. G. Extinction of conditioned psycho-galvanic responses following two conditions of reinforcement. *J. exp.Psychol.*, 1940, *27*, 71-75.

Jenkins, W. O. & Stanley, J. C., Jr. Partial reinforcement: A review and critique. *Psychol. Bull.*, 1950, *47*, 193-234.

Lewis, D. J. Partial reinforcement: A selective review of the literature since 1950. *Psychol. Bull.*, 1960, *57*, 1-28.

Scharf, Eugenia S. The effect of reinforcement in programmed learning sequences. Unpublished master's thesis, University of Pittsburgh, 1961.

Sheffield, Virginia F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, *39*, 511-526.

Skinner, B. F. The science of learning and the art of teaching. *Harvard educ. Rev.*, 1954, *24*, 86-97.

Chapter 5

A Comparison of Three Methods of Programming Subject Material for Auto-Instruction[1]

The purpose of this study was to investigate the effects of three variations in method of controlling the learners' responses to programmed learning material. The effects of these variations in method of controlling the learners' responses were evaluated in immediate and delayed performance measures as well as in the relative efficiency of the three variations as measured by learning time for each learning condition.

Starting with differing views of the learning process, somewhat different methods of programming verbal learning sequences have been suggested by various researchers. For Skinner (1954; 1957; 1958) and others who advocate programming by small steps, channeling of the response is necessary in order to get the learner to respond appropriately in the presence of discriminative stimuli so that the correct response can be reinforced by confirmation. All small-step programmers agree that powerful control over the learner's responses is maintained by the proper sequencing of the material in short, easy steps so that intra-frame prompting can occur, and so that each terminally desired response is more closely approximated as the program progresses. The reinforcement has been primarily, but not exclusively (Silberman, 1960), viewed as resulting from immediate "knowledge of results" (confirmation) occurring at each step.

Skinner-type programs emphasize reinforcement as a central process for the acquisition of behavior; behavior is acquired as a result of a contingent relationship between the response of an organism and the consequent event. The method of small-step programming is believed to meet the contingencies for reinforcement, i.e., the reinforcing event (confirmation) closely follows the occurrence of the behavior, and a large number of reinforcements are given through the use of short, easy steps under a variety of stimulus conditions so that the future probability of occurrence of the appropriate response is high for varied discriminative stimuli. The techniques of small-step programming have been formally expressed by Glaser (1960).

---

[1] The data presented in this report are abstracted from a dissertation submitted by Robert B. Hessert to the Graduate School, Division of Natural Sciences, University of Pittsburgh, (October 1961), in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Details of the study can be obtained from the complete thesis.

The method of programming described as Skinner-type programming uses prompted responding and reinforcement of successive approximations to the terminal responses in sequentially ordered small steps. In this study it is referred to as the prompted and reinforced responding method (PR).

Some recent investigations by Cook and Kendler (1956), Cook (1958), Cook and Spitzer (1960), Irion and Briggs (1957) and others suggest that an effective means of controlling the response is by directly prompting the response with the desired answer before the learner is required to respond. Kopstein and Roshal (1955) indicate that presenting both the stimulus and response members together before requiring a response to the stimulus alone may prove even more effective in eliciting the desired response under subsequent testing conditions without prompting.

This method of prompting with the correct answer prior to requiring a response has not been applied to tasks of the verbal learning sequence type. However, it can be assumed that Cook's method also would employ sequentially ordered, small-step successive approximations to the terminally desired responses. The prompting method suggested by Cook does not require the use of auxiliary stimuli for prompting purposes since the desired response (answer) is given in the presence of the discriminative stimuli prior to the elicitation of the response from the learner. Under this condition, reinforcement (confirmation) of the learner's response is viewed as unnecessary. Cook-type programs have yet to be built, but central to their construction would be Cook's view of the learning process that includes mediating responses (which have stimulus properties) that are made to both the stimulus and response terms. Cook's view appears to be that program construction methods should be modeled on a classical conditioning paradigm rather than on an operant conditioning paradigm.

The Cook method uses non-confirmed responding prompted by the prior presentation of the desired S-R connections in sequentially ordered, small-step approximations to the terminal responses. In this study this method is referred to as the correct answer prompted responding method (AP).

In the work of Pressey (1926; 1959; 1960), the learner is viewed as a selective, self-organizing mechanism. The practical implications of this view are apparent in the freedom allowed the programmer in preparing the program. Since detailed small-step programming is not considered necessary, more of the burden for appropriate selective responding is placed upon the learner, and the learner is given more degrees of freedom in making his responses.

Pressey states that his methods function in accordance with some earlier views of the learning process (Thorndike, 1913). He points out that the "law of recency" operates to establish the correct response, since the last answer chosen is the right answer. The correct answer must inevitably be the most frequently practiced answer because a wrong response, when made, must immediately be compensated for by a further correct reaction (law of exercise). Since the learner can progress only by making the right reaction, he is penalized every time he responds incorrectly by being required to answer the question once more, and is rewarded for right responses by the elimination of that question. Therefore, according to Pressey the "law of effect" is constantly operating to further the learning.

Pressey's method of reinforced practice on test items interspersed throughout periods of free responding to study materials is predicated upon his view that the learner is able to proceed on his own with proper guidance. This guidance for the learner is provided by immediate knowledge of results and some indicators as to what to study if the response is incorrect. In his later studies, the tests are taken under the reinforced retained condition, in which the correct answer must be obtained on each item before proceeding. In addition, the learner keeps his test to guide his further study until the next practice test is taken. Pressey (1960) recognizes that small-step programmers will disagree with him on at least three points: (a) supposedly, very easy learning-producing items are not included; (b) the elimination of substantial numbers of easy items leaves only the more difficult items and hence the program becomes a series of somewhat disconnected probing questions or problems; and (c) learning seems to result from mistakes--immediately corrected.

Thus, Pressey advocates reinforced practice of terminal responding interspersed throughout the transitional phases of learning as guidance for free responding to non-sequenced study materials. In this study this method is referred to as the guided and reinforced responding method (GR).

Procedure

The learning task used in this investigation was a symbolic logic task adapted from the one developed by Evans (Chapter 3). The adaptation of the symbolic logic program used in this study was called the Deductive-Proof Program (DPP). The DPP was developed in four sections. Essentially, Section I consists of a "warm-up" section that is designed to establish a relevant repertoire of responses prior to any experimental variations which might be used in later sections of the program. Section II of the DPP teaches ten rules of symbolic operation drawn from the calculus of propositions. Section III of the program consists of two sets of ten deductive-proof problems, one on each of two levels of complexity. These are used to give the Ss practice on samples of the kinds of tasks that they are required to perform in the criterion test. Section III can be sequenced to follow Section II, or it can be integrated with it. In this study, Section III was integrated with Section II in all three methods: PR, GR, and AP. Section IV is the deductive-proof criterion measure in two 30-item forms, Form A and Form B.

Experimental Treatments

The experimental variations of the combined Section II and Section III of the DPP are explained below.

Prompted and Reinforced Responding Condition: Treatment PR. The "Ruleg-type program" and its method of construction has been described in detail in Chapter 2. This method of programming was the Prompted and Reinforced (PR) Treatment in this study. The number of frames and the sequencing of the frames for this section was determined by empirical tryout with high-school subjects. The overall ordering of Section II frames was determined by the Section III deductive proofs that immediately followed. For example, the first deductive proof item involved the use of two logical rules and the sequence began by having the subjects respond to incomplete rules and incomplete examples for these two rules. These responses were prompted by both completed rules and examples, and by the sequencing of the frames. The Ss responded with the aid of prompts and had their responses confirmed with the answer on the back of each frame. This method was devised following the methods of programming advocated by Skinner and his associates.

Correct Answer Prompted Responding Condition: Treatment AP. The AP method of programming was adapted from the PR method. In the PR Treatment items were ordered on each card with the completed rules and examples used as prompts for the incomplete rules and examples. However, in the AP Treatment each item was responded to by the Ss. While in the PR Treatment a frame or card contained as many as four items, in the AP Treatment there was only one item per frame. The sequencing used in the PR Treatment was maintained in the AP Treatment. The method of presentation in the AP Treatment was to give, on one side of a card, the complete rule statement or example (S-R), and, on the reverse side of the card, the incomplete rule statement or example (S-). This method or programming was devised to represent Cook and Kendler's view of the learner.

Guided and Reinforced Responding Condition: Treatment GR. In the GR Treatment, reinforced practice was given with Section III deductive-proof frames. These were the same sets of items used in the PR and AP Treatments, with the exception that no prompting or correct answer was given. These deductive-proof problems were taken in the retained condition, i.e., the Ss responded, checked the answer side, and if incorrect, either tried the problem again or looked at the study sheets which were available. After responding correctly they then proceeded to the next item.

There were several study sheets for each of the ten rules. These sheets listed the statements of the ten rules and all examples for each rule. These rules and examples were the same ones used in the other two treatments. However, in the other two treatments some items in this section of the program were incomplete; in the GR Treatment all of these items were presented in their completed form.

For the GR Treatment the sequence of events was as follows:

1. All study sheets were handed out to the Ss.
2. Subjects were given instructions regarding the use of each sheet and informed that they had 30 minutes to study before beginning the first practice test of deductive-proof problems.
3. After completing the first (low level of complexity) practice test, the Ss turned it in to the experimenter for a second (higher level) practice test.
4. All three sections were allowed to proceed at their own pace. The development of the GR method of programming follows the suggestions of Pressey and his associates.

Criterion Measures

The major dependent variables for this study were immediate and delayed retention as measured by two forms of a 30-item deductive-proof task. The tests were called the Deductive-Proof Tests, Form A and Form B. Evans, Glaser, & Homme (Chapter 3) and Stillman (Chapter 4) used similar tests in their studies and found them highly reliable.

The two forms of the test for immediate and for delayed recall testing contained similar items. The equating was done by matching the rule combinations and levels of complexity item by item. Each set of matched items from the two tests differed only in the elements used and, in some instances, in the order in which the two rules were applied to reach the concluding step.

Subjects and Administration of the Program

A pool of Ss for this study was selected from the Junior Class of Mt. Lebanon High School, Pittsburgh, Pennsylvania. The selection of 96 Ss was made using a table of random numbers and limiting the population to Ss with I.Q.'s between 100 and 130. This procedure was followed in order to get above the "floor" established in Stillman's study (Chapter 4) which used a program very similar to the PR condition in this study. She found that students with I.Q.'s below 100 had a difficult time in following instructions, comprehending the task, etc. Preliminary work with the DPP demonstrated that with self-paced responding the slower Ss (usually the lower I.Q.'s) took as long as six hours (about one school day) to complete the program. Subjects above 130 I.Q. were also excluded from the pool in order to further reduce within-group variance. The measure of I.Q. used in the present study was taken from high school records obtained during the students' current academic year using the Otis Self-Administering Tests of Mental Ability, Higher Examination.[2]

Because nine of the original 96 Ss in the pool declined when asked to participate, 87 Ss remained for random assignment to the experimental conditions. The study in Chapter 3 reported no differences between the sexes; therefore, both sexes comprised the group of subjects assigned to the experimental treatments in this study with no matching by sex attempted.

The administration of the program was as follows:

1. A common set of instructions was given for the standard Section I of the program. This section was administered in groups of 20 subjects.
2. For the experimentally varied Sections (Sections II and III of the DPP) the Ss were divided into groups of ten each, with one proctor for each group assigned to separate rooms. All Ss were carefully observed to insure that they were following the appropriate instructions.
3. One experimental group was run each day with separate instructions for each experimental condition. The deductive-proof frames in Section III were the only frames responded to overtly, and the Ss were given answer pads on which to record their answers. The proctors observed the S's responses to insure no items were skipped. In all three conditions the Ss were self-paced. The total time required for each program was between four and five hours with short rest pauses about every 90 minutes.

---

[2] Otis, A. S. Otis Self-Administering Tests of Mental Ability, Higher Examination. New York: World Book Company, 1928

4. The criterion measure was administered simultaneously to all Ss within each group. All Ss were given at least a five minute rest period before beginning the criterion test. In order to discourage any tendency on the part of the Ss to rush to the completion of the task, no one was allowed to leave the room.
5. The immediate retention test was administered with a one minute and 50 second time limit on each item.
6. The unannounced delayed retention test was administered two weeks later to each group, also with a one minute and 50 second time limit per item.

Because of the abstract nature of the task and observations made in Stillman's study in Chapter 4, extra-program incentives were applied. A means of conditional payment for correct responding to the Section III and criterion items was established. Each subject was informed that certain pay-items were included in these items. The items were not marked and they were both easy and difficult items. The rate of pay was $ .25 per pay-item. The subjects were informed regarding the amount of total earnings which could be earned by any one subject. Observation of the Ss during the experimental sessions indicated that they were intent on the task and that the incentive used appeared to be effective in gaining the S's best effort. All Ss assigned to the three experimental conditions completed the task, including the retention test administered two weeks following the experimental sessions.

Results

To recapitulate, the objective of this study was to investigate the effects of three variations in method of controlling the learners' responses to programmed learning material. The three methods have been described and given descriptive labels. The method proposed by Skinner and elaborated by Evans, Glaser and Homme (1960) at the University of Pittsburgh as the "Ruleg System of Programming" has been called the Prompted and Reinforced Responding Condition (PR) in this study. The variation in programming following some of the suggestions of Pressey and his associates has been labeled the Guided and Reinforced Responding Condition (GR). The method devised from Cook and Kendler's view of the learner has been called, in this study, the Correct Answer Prompted Responding Condition (AP).

Certain controls were imposed and assumptions made in the experimental design, development of the program variations, and assignment of Ss to the treatments:

1. As previously indicated, the written mode of response was used for all program variations in Section III (deductive-proof items). These deductive-proof items were integrated into Section II of the DPP, which was also experimentally varied according to the three programming methods. However, the responses to Section II items were made covertly by the Ss. Thus, the three treatments -- PR, GR, and AP -- were equated for the response mode, and just the responses to the completion type, deductive-proof items were scored for comparison of errors under the three treatments. All Ss responded to the same items and responded covertly or overtly to the same items in all three treatments.

2. Learning Errors. One of the major assumptions in this study is that the programming methods described above are three variations in the degree of control exerted over the learners' responses. All failures on the part of the Ss to respond correctly to the practice-test items of Section III were viewed as instances of inappropriate or inadequate control over the learners' responses. These instances were scored as learning errors.

An analysis of variance of the Ss' learning-error scores in the three treatments resulted in the rejection of the null hypothesis. The differences obtained between the mean error scores for the three treatments are statistically significant at the .01 level. Thus, these were differential effects on the programming methods on the error rate, with a greater mean error rate for the Guided and Reinforced Responding Condition over the Prompted and Reinforced Responding Condition and over the Answer Prompted Responding Condition in this study. However, individual t tests between treatments showed only a highly significant difference ($P < .01$) between GR and the other conditions, even with degrees of freedom halved because of the significant F ratio resulting from heterogeneity of variance (Edwards, 1950). The mean difference in error rate between the Answer Prompted Responding Condition and the Prompted and Reinforced Responding Condition was not significant when evaluated by the t test, ($P < .05$). Hence, the assumption concerning the differential effects of the three methods of programming on the program learning-error rates is acceptable only in part. Only the GR Condition was significently different in its mean error rate from the PR and AP Conditions.

3. A fundamental consideration in the evaluation of the above reported difference in program learning-error rates for the three treatment groups in their initial ability. As indicated, the assignment of Ss to the three experimental treatments was made according to random selection from a pool of Ss from the eleventh grade at Mt. Lebanon High School with I.Q.'s ranging from 100 to 130. The difference between the mean I.Q. of the treatment groups when evaluated by an analysis of variance was not significant ($P < .05$). It can be assumed that these Ss with respect to their I.Q.'s were from a common population of eleventh grade students at Mt. Lebanon High School. Furthermore, an analysis of learning time and error scores on the standard learning task, Section I of the DPP, indicated common learning ability for the three treatments. Neither the F ratio obtained for the error scores nor the one obtained for the time scores was significant ($P > .05$). From the latter two analyses it can be assumed that the three experimental groups were drawn from a relatively homogeneous population with respect to learning ability, as demonstrated with the materials and procedures of the standard Section I of the DPP.

Analysis of Immediate and Retention Performance Scores

In this analysis the S's performance was evaluated on both the immediate post-program test (Form A) and the retention test administered two weeks later (Form B). The S's performance was scored for errors. By treating the immediate and retention tests as separate trials, a repeated-measures analysis of variance (Edwards, 1950) can be performed on the scores. Such a design permits three sources of variation to be tested for significance: (a) differences due to experimental treatments; (b) differences between trials (immediate performance versus retention); and (c) interaction of treatment and trials. Table 5.1 presents the repeated-measures analyses of variance on the error scores for the immediate and retention tests.

Table 5.1

Analysis of Variance of Error Scores on the Immediate and Retention Tests used as Separate Trials under Three Experimental Treatments

Source of Variation	df	Mean Square	F	P
Between treatments: (PR), (GR), (AP)	2	170.81	1.44	NS >.05
Between Ss in same group	84	113.36		
Total between Ss	86			
Between tests: A, B	1	33.20	2.67	NS >.05
Interaction: tests x treatments	2	9.70	.78	NS >.05
Interaction: pooled Ss x tests	84	12.43		
Total within Ss	87			
Total	173			

With respect to the effects of the treatments, it appears that experimental treatments had little effect on the error score, either as a primary source of variation or in combination with tests x treatments interaction. The fact that the number of errors did not differ significantly over a two-week retention period agrees with the findings of Chapter 3, where the deductive-proof test showed no significant change in errors over a one-week retention period.

Analysis of Learning-Time Scores on the Experimental Treatments

As previously discussed, it was the second and third sections of the DPP that were varied according to the three programming methods and comprised the experimental treatments. The Ss in each treatment group responded at their own pace. The mean times for each experimental treatment group should provide an index of the relative efficiency of learning by the three programming methods.

Table 5.2 presents an analysis of variance of the learning time in the experimentally varied sections of the learning programs (Sections II and III for PR, GR, and AP Treatments). For two and 84 degrees of freedom, an F ratio of 4.68 is required for significance at the .01 level. The differences between treatments are highly significant ($P < .01$). Thus, the differential effects of the programming methods on the learning time has resulted in a significant difference between the means of all the experimental groups. The three t tests are all highly significant ($P < .01$), even with degrees of freedom halved because Barlett's test for homogeneity of variance (Edwards, 1950) proved significant ($P < .05$, but $P > .01$). Thus, the different programming methods required different mean amounts of learning time for the three experimental treatments. However, there was a great deal of variation within groups and overlapping of times between groups. If the criterion of efficiency is applied as measured by S's learning time, the Prompted and Reinforced Treatment appears to be the most efficient. The PR Treatment Ss were, on the average, over 11 minutes faster than the GR Treatment Ss, who were nine minutes faster than the AP Treatment Ss. Thus, the mean savings in learning time for the PR Treatment Ss was nearly 22 percent when compared with the AP Treatment Ss; when the PR Treatment was compared with the GR Treatment the mean savings was almost 13 percent in favor of the PR Treatment Ss.

Table 5.2

Analysis of Variance of Learning Time on the Three Experimentally Varied Sections of the Learning Task, Sections II and III of the Deductive-Proof Program

Source of variation	df	Mean Square	F	P
Treatments	2	3103.82	36.78	< .01
Error	84	84.39		
Total	86			

The Reliability of the Criterion Measures

Reliability was estimated from the results obtained for two parallel test forms (Forms A and B) administered two weeks apart. The correlation coefficient obtained of $r = .82$ was for the combined treatments. When the somewhat restricted range of ability of the groups and the length of the test are also taken into consideration, this estimate of reliability must be considered conservative.

Discussion

One general statement can be made concerning the present study. With respect to statistical significance obtained, the effect of the experimental treatments was on time to learn by the different programming methods and not on criterion performance. Treatment effects were present on time-to-learn scores and on error rate while learning. However, treatment effects on criterion performance did not approach significance.

Reviewing the results, and keeping in mind that the three treatment groups were essentially equivalent in initial performance on a standard task, the effects of the treatments on learning performance and learning time must be considered. The Guided and Reinforced Responding (GR) Treatment Ss made over three times the number of errors made by the Answer Prompted Responding (AP) Treatment Ss, and over twice the number of errors made by the Prompted and Reinforced Responding (PR) Treatment Ss on the experimentally varied portion of the program. It appears by these results that the three treatments did exert different degrees of control over the learners' responses, as assumed. However, the crucial question to answer is why this difference in degree of control over the learners' responses did not result in differential effects on the criterion performance. The experimental design does not provide an unambiguous answer to this question; however, the results of several other studies, using similar programmed materials, can be used to reveal more about this question.

First, in the present study, it is possible that the differences in programming methods may have provided compensating factors that would cancel the differential effects of the learning-error rates under the different treatments. The PR Treatment provided a high degree of intra-and inter-frame prompting plus immediate knowledge of results by providing the correct answer on the reverse side of each card. Thus, incorrect responses could be immediately corrected, but they were not called for again in the same form. Correct responses were immediately confirmed. The AP Treatment provided for a high degree of intra- and inter-frame prompting by following the same intra-frame prompting sequence as in the PR Treatment plus prompting a specific response to every item by prompting with the correct answer; however, in this case, no knowledge of results was provided after the response. The GR Treatment Ss had no programmed sequence of prompting, but responded to the practice test items in the retained position until they obtained a correct response. The correct answer was provided on the reverse side and the Ss made use of study guides to aid them if they made an incorrect response. Thus, in this treatment the Ss responded with immediate knowledge of results. They could immediately correct their incorrect responses, and, in fact, they had been instructed to correct them before continuing. Their correct responses were immediately confirmed.

The Ss in each group were exposed to the same material, but they responded under varying degrees of prompting. The highly prompted AP Treatment Ss, however, did not have their responses confirmed. The low-prompting GR Treatment Ss had all responses confirmed, and were instructed to continue responding until confirmed as correct. The PR Treatment fell in between the other two treatments in amount of prompting, but it also provided for immediate knowledge of results for each response without requiring the Ss to continue responding until correct.

The fact that more errors were made in the GR Treatment may have been compensated for by immediate knowledge of results and repeated responding until the correct responses were made. Thus this method conformed with Pressey's view of the learner and the learning theories to which he subscribes. The last response was the correct response and was reinforced, while an incorrect response was punished, i.e., it had to be corrected. Pressey views the learner as an adaptable, self-correcting organism, one who can "profit" from mistakes if given immediate knowledge about the results of his responses and about how to correct them if they are in error.

The AP Treatment, with its answer prompting and small-step sequences, was not expected to result in many learning errors. This method of programming did not provide immediate knowledge of results. However, the question was whether or not it would be as effective as the other two methods for criterion performance. The method proved to be as effective as the other two methods at criterion performance. Confirmation of the responses was not necessary with this method, as Cook (1958) has demonstrated in another learning context.

The PR and AP Treatments did not differ statistically in their mean error rates. A careful scrutiny of the two methods reveals that they do not differ greatly in their basic construction. They are both highly prompted sequences. Intra-frame prompting in the AP method is done by using the correct answer as the prompt for the following frame (S-R, then S with the student responding with R). In the PR method there is some intra-frame prompting, although additional intra-frame prompting is provided by analogy (from an example) and/or by deduction (from a rule statement). In both instances the probability of a correct response is undoubtedly very high; therefore, confirmation of the response is most likely unnecessary for learning.

In Stillman's (Chapter 4) study, where a highly prompted sequence was used, she found no differences in criterion performance with four treatments composed of four different proportions of answers on the reverse side of the frames. She concluded that different ratios of reinforcement with the correct answer did not demonstrate any treatment effects on criterion performance with a highly prompted program. Her program, in content and form, very closely resembled the PR Treatment in the present study.

Chapter 3 has already contrasted the differences between the classical techniques of investigating learning in a low probability context, where extraneous cues, associations, etc., are carefully controlled and in the programmed sequences, where the probability of obtaining the correct response is enhanced by the use of prompting devices. Evans (Chapter 3, page 73) concluded that "the relevance of variables such as response mode and immediacy of confirmation was inversely related to the probability of correct responding." Stillman's findings would modify this statement to include that the effects of variations in the ratio of confirmation have reduced influence when the probability of correct responding is high. However, she did not specifically test the inverse relationship. The present study more closely approximates a test of this relationship. It provided one treatment (AP) with no confirmation in a high probability of correct responding context, and another treatment (GR) with very little direct prompting and a high rate of response errors which provided immediate feedback and continued responding until the correct response was confirmed. However, a balanced experimental design with additional treatments as controls would be necessary before the inverse nature of this relationship could be properly evaluated. This would entail AP type programs with varying ratios of confirmation and programs using the GR method with varying ratios of confirmation of the responses.

Since the three treatments in this study produced, essentially, no differences in either criterion performance or retention, the next relevant question might be: Which was the most efficient method? Although there was a great deal of variability and an overlap in learning time among the three experimental groups, they also differed considerably in the average time to learn by the three methods. These differences were statistically significant and favored the PR Treatment over the GR Treatment, which, in turn, was favored over the AP Treatment. Thus, in this study the combined prompting and confirmation method of programming used in the PR Treatment was the most efficient of the three methods without sacrifice in learning effectiveness.

Summary

Different methods of programming auto-instructional study material have been suggested in the literature. These methods were analyzed and reduced to three variations in the controlling of the learners' responses to programmed material. The purpose of this investigation was to evaluate the effects of these three variations upon immediate and delayed performance measures, as well as on the relative efficiency of the methods as measured by learning time for each condition. The methods of controlling the learners' responses were given descriptive labels according to the combinations of prompting and confirmation used in each method. The three treatments were: the Prompted and Reinforced Responding Treatment; the Guided and Reinforced Responding Treatment; and the Answer Prompted Responding Treatment.

The three treatments were administered to 87 high school subjects randomly assigned to the three experimental conditions. These three independent groups of 29 each were evaluated as having equal initial ability to learn subject material in symbolic logic.

By treating the immediate and retention tests of deductive-proof problem solving as separate trials, a repeated-measures analysis of variance was performed on the error scores for the criterion measures. A separate analysis of variance of learning-time scores was made for the three treatments. With respect to statistictical significance obtained, the effect of the experimental treatments was on time to learn by the different programming methods and not on criterion performance. The Prompted and Reinforced Responding Treatment group was the most efficient of the three treatment groups.

The three treatments produced significantly different learning-error rates by the three experimental groups. The fact that no differential effects in criterion performance were produced by the differences in learning-error rates was explained on the basis of compensating factors present in the three programs. The Answer Prompted Responding Treatment, which was programmed so that no confirmation of response was possible, produced the lowest learning-error rate. This low error rate was made possible through the use of maximum prompting by showing the students in this treatment both the stimulus and the required response before requesting the students to respond to the stimulus alone. The Guided and Reinforced Responding Treatment, which had minimal prompting and produced the highest learning-error rate, provided for immediate confirmation of the response and continued responding until the correct response was confirmed. The Prompted and Reinforced Responding Treatment, which yielded the next highest error rate, provided for immediate confirmation of responses and a substantial amount of prompting. However, the prompting for the latter treatment (PR) was provided by the use of auxiliary stimuli and, therefore, was less effective in reducing error during learning than the prompting in the Answer Prompted Responding Treatment.

The implication of this investigation for programming is discussed in terms of the effect of reinforcement in the high response probability learning situations presented by programmed learning.

References

Cook, J. O. Supplementary report: Processes underlying a single paired-associate item. *J. exp. Psychol.*, 1958, *56*, 455.

Cook, J. O. & Kendler, T. S. A theoretical model to explain some paired-associates learning data. In G. Finch & F. Cameron (Eds.), *Symposium on Air Force human engineering personnel, and training research*. Washington, D.C.: National Academy of Sciences - National Research Council, 1956. Pub. 455. Pp. 90-98.

Cook, J. O. & Spitzer, M. E. Supplementary report: prompting versus confirmation in paired-associate learning. *J. exp. Psychol.*, 1960, *59*, 275-276.

Edwards, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.

Evans, J. L., Glaser, R., & Homme, L. E. *An investigation of teaching machine variables using learning programs in symbolic logic*. A report to the U.S. Office of Education, University of Pittsburgh, November 1960.

Evans, J. L., Homme, L. E., & Glaser, R. *The Ruleg System for the construction of programmed verbal learning sequences*. A report to the U.S. Office of Education, University of Pittsburgh, August 1960.

Glaser, R. *Principles and problems in the preparation of programmed learning sequences*. A report to the U.S. Office of Education, University of Pittsburgh, September 1960.

Irion, A. L. & Briggs, L. J. *Learning task and mode of operation variables in the use of the subject matter trainer*. Lackland AFB, Air Force Personnel and Training Research Center, October 1957. AFPTRC TR 57-8.

Kopstein, F. F. & Roshal, S. N. Methods of presenting word pairs as a factor in foreign vocabulary learning. *Amer. Psychologist*, 1955, *10*, 354. (abstract)

Pressey, S. L. A simple apparatus that gives tests and scores--and teaches. *Sch. & Soc.*, 1926, *23*, 373-376.

Pressey, S. L. Certain major psycho-educational issues appearing in the Conference on Teaching Machines. In E. Galanter (Ed.), *Automatic teaching: The state of the art*. New York: Wiley, 1959. Pp. 187-198.

Pressey, S. L. Some perspectives and major problems regarding "teaching machines." In A. A. Lumsdaine & R. Glaser (Eds.), *Teaching machines and programmed learning: A source book*. Washington, D.C.: National Education Association, 1960. Pp. 497-505.

Skinner, B. F. The science of learning and the art of teaching. *Harvard educ. Rev.*, 1954, *24*, 86-97.

Skinner, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.

Skinner, B. F. Teaching machines. *Science*, 1958, *128*, 969-977.

Silverman, R. E. *Automated teaching: A review of theory and research.* Port Washington, New York, June 1960. Technical Report NAVTRA DEVCEN 507-2.

Thorndike, E. D. *Educational psychology. Vol. II: The psychology of learning.* New York: Teachers College, Columbia University, 1913.

Chapter 6

The Application of Programmed Learning Techniques to Selected Material in a Professional Curriculum
A Study of Retention Effects[1]

Post-graduate professional education currently faces increasing demands with respect to the equality and quantity of graduates, a trend which promises to increase in severity in the near future. Professional schools will be asked to increase their output of graduates, probably without proportional increases in materials and manpower. It seems unlikely that any appreciable relief will result by urging greater effort from professional faculties whose members already work at or near capacity. If anything, a lessening in graduate quality would be the result of such an increase in student load per faculty member. Recent developments in the field of programmed learning suggest a solution which certainly merits at least a tryout at the professional level. No appreciable effort to introduce programmed learning into a professional curriculum has yet been reported. One of the chief difficulties in such an effort is the complexity of the material dealt with and the scarcity of qualified and interested programmers. This paper reports the results of an application of programmed learning instruction in a relatively complex subject matter that is part of the training given in the University of Pittsburgh School of Pharmacy. The experiment reported here compared programmed and non-programmed subject matter presentation and provided comparative tests of immediate and delayed recall of the subject matter.

The subject material selected for programming was a "knowledge and understanding of vitamins." The format used consisted of a variation of the programmed textbook. This mode of presentation provides a workbook so arranged that immediate confirmation is provided by turning the page. The programmed textbook has been found useful in a number of situations where the maturity of the student or the nature of the program permits the assumption that the student will not "peek" at the answer before trying the item. The programmed text further permits the elimination of costly machines.

---

[1] This report is a version of a paper originally submitted by Carole M. Finelli to the Graduate Faculty, Division of Natural Sciences, University of Pittsburgh (1961) in partial fulfillment of the requirements for the degree of Master of Science. Dr. Robert W. Sager, Professor of Pharmacy, was a major advisor in this work.

Materials and Procedures

Characteristics of the Programs. The material taught in the course of this study was limited to the field of vitamins, a field which constitutes a specific section of a professional course in pharmacy given during the terminal year in the School of Pharmacy at the University of Pittsburgh. The type of programming used in this study was of the "Skinnerian" variety. That is, frames were short, the answers filled in by the students were short and specific, and immediate confirmation of the answer was provided. The technical information about vitamins which was taught by the programs was gathered from standard references in the field: Textbook of Biochemistry by Harrow and Mazier (1959) and Biochemistry by Cantarow and Schepartz (1957) as well as from material from recent undergraduate and graduate lectures. Because of the exploratory nature of this attempt to place programmed learning into the professional curriculum, no hard and fast procedure was adapted prior to the start of the experiment; the programming and presentation of the material selected proceeded according to feedback from the students being used as subjects.

Program Construction. Vitamin A was the first to be programmed. It was arranged in booklet form, the size being approximately eight by three inches (see Figure 6.1).This programmed booklet contained the frame on one page, and the correct response to the frame on the next page. The second frame appeared adjacent to the response of the first frame. This procedure was followed throughout the Vitamin A program, that is, the response to one frame appeared on the same page with the succeeding frame. Each student had his own booklet, and wrote his answer to an item on a separate sheet of paper. He then turned the page of the booklet and compared his answer with the correct response. If he had made an incorrect response, he so indicated with a check mark.

Figure 6.1

Examples of the first two pages of the Vitamin A booklet

Page One of the Vitamin A Booklet

Introduction to Vitamin A

1. The first vitamin to be studied derives its alphabetical designation from the fact that it was the first substance to be placed in the category of nutritionally essential factors. This vitamin is known today as ______________ ____.

Page Two of the Vitamin A Booklet

---

1. Vitamin A	2. Vitamins are called "accessory food factors," and are necessary for normal nutrition and growth. When the needed amount of a vitamin is not present in the body, there is a vitamin ______________.

It was found that the appearance of the response to one item on the same page as the succeeding item sometimes distracted the student, and often supplied an unwanted prompt or cue. For this reason, it was decided to place the response to an item on a separate page of its own. The second program prepared was that of Vitamin D. The size of the booklet was the same as that of Vitamin A, and the only change was the doubling of the total number of pages so that separate pages were provided for responses (see Figure 6.2).

Figure 6.2

Examples of the first two pages of the Vitamin D booklet

Page One of the Vitamin D Booklet

---

1. The inactive natural precursors, or provitamins, of the D vitamins are cyclopentanophenanthrene derivatives. They are therefore classified as steroids. Thus, Vitamin D is the only vitamin to have a ____________ structure.

Page Two of the Vitamin D Booklet

---

2. steroidal

Experience with the Vitamin D program led to the opinion that two further modifications were required. First, while the response to an item had been removed as a source of prompting for the succeeding item, the type of white mimeograph paper used was sufficiently translucent to allow the response to be seen through the paper. Second, it was noted that the size of the booklet made handling of it somewhat awkward. To remedy these conditions, less translucent colored paper was used in subsequent programs. The 3" x 8" format was also dropped, and subsequent booklets were prepared on standard 8½" x 11" blue opaque mimeograph paper (see Figure 6.3).

Figure 6.3

An example of the first and second pages from the Vitamin B_{12} booklet

Page One of the Vitamin B_{12} Booklet

---

1. Introduction

There is a group of complex compounds possessing Vitamin B_{12} activity. These compounds are members of the B-complex group, and are __________-soluble vitamins.

---

11. Occurrence

Substances with Vitamin B_{12} activity are present in liver, eggs, milk, meats, and fish. Minute amounts are probably present in all animal cells. Vitamin B_{12} does not occur in plant tissues. Green vegetables __________ (would or would not) be a good source of Vitamin B_{12}.

---

21. 5. Absorption of Vitamin B_{12} is facilitated by the presence of the __________________.

6. The "intrinsic factor" is a polypeptide or ________ present in __________________.

7. Without the presence of the "intrinsic factor," Vitamin B_{12} ______________________________.

---

31. Deficiency of Vitamin B_{12} results in a macrocytic type of anemia. From the synonym "anti-pernicious anemia," it is evident that Vitamin B_{12} functions in the prevention of __________ __________.

---

41. Requirement

Although large amounts of Vitamin B_{12} are synthesized by intestinal bacteria, probably very little if any is available to man from this source because it is produced largely in the colon, where its __________ is highly questionable.

Figure 6.3(continued)

Page Two of the Vitamin B_{12} Booklet

---

1. water

---

11. would not

21.
5. intrinsic factor
6. protein
normal gastric juice and saliva
7. would not be absorbed (from the intestines)

---

31. pernicious anemia

---

41. absorption

Following administration of the programs for Vitamins A and D, the remaining two fat-soluble vitamins, E and K, were presented to the class by the traditional lecture method. Following this, the thiamine and riboflavin programs were administered to the class. An examination was prepared covering the six vitamins: A, D, E, K, thiamine, and riboflavin. (Criterion measures are discussed below.) After the students had been given instruction in six of a total of 13 vitamins--two by lecture and four by program--the first examination was administered. This marked the completion of the major experimental phase. During the remainder of the course in which the subjects were enrolled, other vitamins were programmed and taught in this fashion. The vitamins programmed following the first examination were: Vitamin C, pyridoxine, niacin, pantothenic acid, biotin, folic acid, and Vitamin B_{12}.

In an attempt to measure the degree of retention of material on the first six vitamins, four of which were programmed and two given by lecture method, the same examination given earlier was repeated, unannounced, six weeks later. The students were informed at the beginning of the testing period that they were being subjected to the re-examination for research purposes only, and the results would not affect their grade in the course.

After all 13 vitamins had been administered, an examination was prepared on the latter seven vitamins, which had not been covered in the previous examination. A portion of the final examination in the total course also consisted of 50 matching type questions covering all 13 of the vitamins. At the time of the final examination in the course, the students were given a questionnaire to determine their reactions to programmed instruction.

Criterion Measures

Examination I (covering vitamins A, D, E, K, thiamine, and riboflavin) consisted of 63 multiple choice questions having five answer choices each. Most questions had more than one correct answer, students being permitted to check as many choices as they thought correct. Scoring was done in terms of total number of mistakes, i.e., checking incorrect choices and failing to check correct choices. Examples of the type of question used in Examination I appear in Figure 6.4.Omitting three ambiguous responses, there remained a total of 312 possible answers. In addition to the multiple-choice questions on Examination I there were 25 association (analogy)questions of which 24 were considered valid questions. Examples of these analogy questions are shown in Figure 6.5. Examination I, as previously indicated, was readministered six weeks following completion of instruction in the first six vitamins as a measure of retention.

After all 13 vitamins had been administered, an examination (Examination II) was prepared on the last seven vitamins taught through programmed instruction. This examination consisted of 30 multiple-choice items each having only one right answer, plus 20 association questions. A final facet of the evaluation procedures was a part of the final course examination which consisted of 50 matching type questions covering all 13 of the vitamins.[2]

---

[2] The association items included in the two administrations of Examination I dealt with more than one vitamin each, that is, the same question might deal with one vitamin presented by program and one presented by lecture. For this reason, only the multiple-choice, multiple-response items were entered into the analysis described below. Although the additional testing, Examination II and the final course examination have been described for the sake of completeness, no detailed analysis of the results of these additional tests will be presented.

Figure 6.4

Examples of Multiple Choice, Multiple Response Questions From the Examination Covering Vitamins A, D, E, K, Thiamine and Riboflavin

---

Vitamin D differs from all other vitamins in that it is

(x) steroidal in character.
( ) fat soluble.
( ) involved with rhodopsin.
(x) involved with alkaline phosphatase.
( ) formed from food precursors.

Among the symptoms of Vitamine A deficiency would be

(x) xerophthalmia
(x) hyperkeratosis
(x) nyctalopia
( ) rachitic metaphysis
(x) night blindness

The ordinary, but unproven, requirement for Vitamin K is

(x) one or two mg. of menadione.
( ) one or two mg. of K_1.
( ) five mg. of menadione.
(x) five mg. of K_1.
( ) none of these.

Among good food sources of Vitamin E are

(x) green plants.
(x) vegetable fats.
( ) animal fats.
(x) wheat germ oil.
( ) fresh pork.

Extreme thiamine deficiency can lead to

(x) dry beriberi.
(x) wet beriberi.
(x) edema.
(x) congestive heart failure.
(x) degeneration of nervous tissue.

Figure 6.5

Examples of Association Questions
From the Examination Covering Vitamins A, D, E, K,
Thiamine and Riboflavin

Beriberi : Thiamine :: Rickets : (Vitamin D)

Vitamin K : Prothrombin :: Vitamin A : (Rhodopsin)

Anti-hemorrhagic : Vitamin K :: Antioxidant : (Vitamin E)

Night blindness : Nyctalopia :: Eye dryness : (Xerophthalmia)

5,000 : Vitamin A :: 400 : (Vitamin D)

Calcification : Vitamin D :: Tissue respiration : (Riboflavin)

Subjects. The subjects in this study were 49 University of Pittsburgh pharmacy students who had been exposed almost exclusively to the traditional lecture method of teaching for the previous seven semesters, as well as for their concurrent courses during the study.

Sampling Student Opinion. In addition to the programs and examinations, a questionnaire was prepared and administered to the students during the final examination in the course. This questionnaire was designed to determine student reactions and opinions regarding the programmed textbook method of education. The questionnaire and the results obtained from it are shown in Figure 6.8 which has been included in the Results section of this report.

Figure 6.6

Results of the Examination on Vitamins A and D, E and K

Multiple-choice, Multiple-response Questions
(95 possible answers for Vitamins A and D)
(75 possible answers for Vitamins E and K)

Student Number	Vitamins A and D % of correct responses	Vitamins E and K % of correct responses
1.	89.5	88.
2.	95.8	86.7
3.	92.6	84
4.	88.4	93.3
5.	96.8	88.
6.	96.8	89.3
7.	90.5	74.7
8.	96.8	85.3
9.	93.7	86.7
10.	95.8	93.9
11.	98.3	100.
12.	89.5	80.
13.	93.7	88.
14.	95.8	93.3
15.	89.5	89.3
16.	97.9	90.7
17.	95.8	84.
18.	87.4	85.3
19.	98.9	86.7
20.	93.7	83.3
21.	94.7	80.
22.	90.5	92.
23.	92.6	92.
24.	98.9	82.7
25.	95.8	86.7
26.	89.5	80.
27.	97.9	89.3
28.	96.8	84.
29.	94.7	93.3
30.	97.9	88.
31.	91.6	74.7
32.	97.9	88.
33.	93.7	88.
34.	90.5	80.
35.	91.6	94.7
36.	98.9	82.7

Figure 6.6(continued)

Student Number	Vitamins A and D % of correct responses	Vitamins E and K % of correct responses
37.	93.7	89.3
38.	94.7	92.
39.	95.8	97.3
40.	91.6	85.3
41.	96.8	98.7
42.	89.5	90.7
43.	92.6	86.7
44.	97.9	88.
45.	91.6	93.3
46.	95.8	89.3
47.	93.7	82.7
48.	96.8	93.3

Figure 6.7

Results of the Re-examination on Vitamins A, D, E, and K

Multiple-choice, Multiple-response Questions
(95 possible answers for Vitamins A and D)
(75 possible answers for Vitamins E and K)

Student Number	Vitamins A and D % of correct responses	Vitamins E and K % of correct responses
1.	87.4	68.
2.	89.5	81.3
3.	95.8	78.7
4.	88.4	77.3
5.	92.6	85.3
6.	95.8	80.
7.	88.4	69.3
8.	92.6	85.3
9.	94.7	84.0
10.	96.8	81.3
11.	100.	92.0
12.	78.9	88.7
13.	89.5	81.3
14.	97.9	94.7
15.	86.3	76.0
16.	90.5	85.9
17.	92.6	73.3
18.	85.3	78.
19.	96.8	86.7
20.	95.8	89.3
21.	91.6	81.3
22.	92.6	89.3
23.	91.6	78.7
24.	98.9	78.7
25.	93.7	84.
26.	87.4	62.7
27.	91.6	86.7
28.	92.6	88.7
29.	94.7	86.7
30.	90.5	84.0
31.	88.4	84.0
32.	94.7	88.7
33.	87.4	87.3
34.	88.4	70.7
35.	87.4	81.3
36.	92.6	64.0

Figure 6.7(continued)

Student Number	Vitamins A and D % of correct responses	Vitamins E and K % of correct responses
37.	84.2	85.9
38.	84.2	85.9
39.	90.5	85.9
40.	91.6	77.3
41.	100.	90.7
42.	95.8	93.3
43.	90.5	82.7
44.	97.9	85.9
45.	96.8	90.7
46.	95.8	85.9
47.	96.8	73.3
48.	93.7	81.3

Summary of Materials and Procedures. The following is the order in which materials were prepared and administered to the subjects in this experiment:

1. Vitamin A programmed and printed in 3" x 8" format with answers adjacent to the following frame.
2. Vitamin A program administered to class.
3. Vitamin D programmed and printed in 3" x 8" format with separate pages for showing correct answers.
4. Vitamin D program administered to class.
5. Vitamins E and K presented to class in traditional lecture situation.
6. Thiamine (B_1) programmed and printed in 8½" x 11" format on blue paper. Answers appeared alone on following page in same relative position of frame to which they applied.
7. Thiamine (B_1) program administered to class.
8. Riboflavin (B_2) programmed and printed in 8½" x 11" format on blue paper. Answer arrangement was the same as 6 above.
9. Vitamin B_2 program administered to class.
10. Administration of Examination I on Vitamins A, D, E, K, B_1 and B_2.
11. Vitamin C, pyridoxine, niacin, pantothenic acid, biotin, folic acid, and Vitamin B_{12} programmed and printed in same format as 6 above.
12. Last seven vitamin programs administered to class.
13. Examination I readministered six weeks following the first administration.
14. Administration of Examination II covering vitamins listed in 11 above.
15. Administration of final course examination, part of which was devoted to vitamins, and the opinion questionnaire.

Results

Comments on Program Format. Upon completion of all the programs for the 13 vitamins and their administration to the students, a standard programmed textbook for this subject matter was recommended. As previously described, this textbook consisted of 8½" x 11" colored, opaque paper, with items on one page and responses on the immediately succeeding page. Such a programmed textbook overcame the disadvantages which became apparent during the study. As a result of trying various answer sheets, a standard answer sheet was also recommended in which one column was intended for the student's response, and the other column was reserved for him to write the correct response in those cases where he had made an error. The fact that the student had to write the correct response not only increased the probability of fixing it in the student's repertoire, but also of having the response correctly spelled.

The Results of the Two Administrations of Examination I. For purposes of comparison, and prior to the analysis of results, it was decided to eliminate thiamine and riboflavin responses from a study of the first examination results. This was done because it was felt that the four vitamins, A, D, E, and K, all belonging to the fat-soluable with respect to vitamins, were more closely related type of content and level of difficulty and therefore would be more logical subject matters for comparison. Of the 63 multiple-choice, multiple-response questions on Examination I, 19 items having 95 possible answers pertained to Vitamins A and D, and 15 having 75 possible answers referred to Vitamins E and K. It will be recalled that Vitamins A and D were presented by programmed textbook and Vitamins E and K were presented by the traditional lecture method. Because the number of questions on Vitamins A and D differed from the number on Vitamins E and K, percentages of the questions right were calculated for the A and D combination and for the E and K combination. Figure 6.6 shows each student's percentage of correctly answered questions for the two groups of Vitamins, A and D and E and K. In line with the individual student trend, the average performance percentage of correct responses made on the first administration of Examination I for the A and D questions was 94.20 percent while the average percentage of correct responses for the E and K questions was 87.76. Using these percentages, "t" test between the differences was computed. The significance test used was a that for means of correlated observations. With an N of 48, the value of "t" was significant the .001 level leading to the rejection of the hypothesis of no difference.

When the examination was given unannounced six weeks later, the average percentage of correct responses for the A and D questions was found to be 92.16 and the mean percent right for E and K, the non-programmed vitamins was 82.29 percent. Percentages for the individual students are shown in Figure 6.7. A "t" test for the differences between means of correlated observations was computed on the re-examination data. Retention on the programmed presentation was significantly greater than on the non-programmed presentation. Results were significant beyond the .001 level in favor of programmed presentation. It should be pointed out that retention of programmed material was superior although it had occurred prior to the lecture material and hence should have lost more through forgetting.

Next, an analysis of amount of forgetting between presentation methods was made. On the programmed presentation, Vitamins A and D, there was a drop of 2.04 percent in number of correct responses. On the non-programmed presentation, E and K, there was a drop of 5.47 percent in correct responses. A "t" test of this differential decline between testings was computed which showed significantly less loss in retention for the programmed material than for the non-programmed material. Results were significant at the .01 level.[3]

Results of the Opinion Questionnaire. The results of the questionnaire (see Figure 6.8) which was given at the time of the final course examination showed that 47 out of 48 students thought the program technique was a good teaching method. Twenty-six students felt that they had learned more by the programmed textbook method than they would have learned had the material been covered by the lecture method, 17 felt they learned the same amount, while five felt they learned less. From Figure 6.8 it may be seen that, in general, student opinion favored programmed learning, although many students had reservations and suggestions. Some of the quotations from the students' answers to question six reflect many of the advantages to programmed instruction which authorities in the field have voiced. It was the general opinion of the students that the programmed material should be supplemented by recitation periods and small group seminars, and that programs and lectures should be given alternately.

---

[3] The data from the two administrations of Examination I were also subject to a non-parametric analysis. The Wilcoxon rank test for paired observations was used. An analysis was made of the immediate test differences, differences on the retention test, and differences in total amount lost. The two methods of presentation produced differences at the .001 level in both immediate and retention tests. It was again found that the amount of loss was significantly greater for the non-programmed presentation at the .01 level.

Figure 6.8

Results of the Opinion Questionnaire

1. As a general statement, do you feel the experimental type of teaching to which you were exposed is a good method of teaching?

47 yes   0 no   1 no opinion

2. Do you believe you have learned more, less, or the same as you would have, had the material been covered by lecture method?

26 more   5 less   17 the same

3. As a general rule, do you like to listen to lectures on factual material?

14 yes   18 no   13 indifferent

3 had no opinion

4. In the part of this course on vitamins, should there have been more lectures and fewer programs?

11 more lectures and fewer programs
13 more programs and fewer lectures
22 just as it was
2 no opinion

5. If programs are used primarily, what do you think additional classes should be used for? (Answer one or more if you wish.)

13 formal lectures
21 small-group seminars
11 student reports and student participation
29 recitation periods

6. Do you feel you would rather have been taught the subject of vitamins by lecture method or by programmed techniques?

4 lecture method   36 programmed techniques   6 either

Why?

Those answering "lecture method" felt that because this was something new it was interesting, but they felt it could become boring. Others liked the student-teacher relationship felt in a lecture, and being taught by programmed technique was too impersonal.

Reasons given by those answering programmed technique were that "it required concentration and thought"; "higher level of retention"; "it is more effective and a good learning method"; "it teaches important facts"; "learn while taking the program"; "work at own rate"; "no cramming for exams"; "requires less study"; "it is easier"; "don't sit idle as in a lecture"; and "receive material that might have been missed in lecture."

Figure 6.8(continued)

---

7. Do you think you were more tired, or less tired, after a period of programmed learning than after an equivalent period of learning by lecture?

26 more tired   8 less tired   13 no difference

One mentioned he was bored because of a lack of personal influence.

8. Do you believe programmed learning (under the conditions you had) requires more, less, or the same time in outside study than does learning by lecture?

6 more time   33 less time   9 same time

9. Do you believe your mental attitude and/or physical condition influence your ability or capacity to learn from a program?

29 considerable
16 moderate influence
3 no influence

The Relationship Between Past Grade Average and Course Grade. In order to compare student performance by programmed learning to overall grade average through the previous eight semesters, the cumulative quality point average (grade average) was obtained. The quality point averages reported here are based on a three-point system where 3.00 = straight "A" work, 2.00 = "B" and 1.00 = "C." Using catagories of 1.00 to 1.49, 1.50 to 1.99, and 2.00 to 2.99, the class of 48 students was divided into three groups. Group 1 consisted of 11 students with cumulative quality point averages from 2.00 to 2.84. Group 2, consisted of 17 students, and quality point averages from 1.50 to 1.99. Group 3 contained 20 students with quality point averages from 1.18 to 1.49.

The 11 students in Group 1 had an average numerical grade (taken from Examinations I, II, and the final course examination) of 90.9 with a range of 87 to 98. This indicates that these students would master the material whether it was given by lecture or by programmed learning. The 17 students in Group 2 had a course average of 86.6 ranging from 75 to 94. Nine of the students in Group 2 had averages from 89 to 94, four had averages from 83 to 86, and the remaining four had averages from 75 to 58. It may be noted that these students, with cumulative quality point averages equivalent to numerical grades of 75 to 80 performed for the most part well above their usual average.

The 20 students in Group C had an average of 83.3 ranging from 73 to 91. Six had averages from 89 to 91. Nine had averages from 81 to 87, and five had averages ranging from 73 to 77. It is of interest to note that these 20 students, representing nearly half the class, and having cumulative quality point averages equivalent to numerical grades of 72 to 75, performed considerably above their usual numerical standard. Three-fourths of the groups produced grades of 81 or better, which would ordinarily fall into quality point averages above 2.0.

These results of the attempt to relate customary level of performance to performance in a programmed course indicate that learning by programmed textbook, while no more effective for students with higher grades achieved by traditional methods of teaching, may prove useful for those "average" students who comprise the majority of American college classes.

Summary

Programmed-textbook learning was applied to the study of vitamins. Thirteen vitamins were programmed. The programs were administered to 48 University of Pittsburgh senior pharmacy students. Two vitamins were presented by traditional lecture method and achievement was compared with two similar vitamins presented by programmed textbook technique. Three examinations were prepared and administered to determine the degree of the student's mastery of the material. The first examination was repeated unannounced after an interval of six weeks in order to measure retention of the material.

Results of the experimental variations showed significant differences between the programmed and non-programmed presentation of the vitamin programs both for immediate and retained knowledge. Student performance after programmed instruction was generally superior to the performance level that would be predicted by the student's past performance.

References

Cantarow, A. & Schepartz, B. Biochemistry. (2nd ed.) Philadelphia: W. B. Saunders, 1957.

Harrow, B. & Mazier, A. Textbook of biochemistry. (7th ed.) Philadelphia: W. B. Saunders, 1959.

PRECEDING PAGE BLANK- NOT FILMED

Chapter 7

An Investigation of Visual Versus Auditory Programming in Teaching Foreign Language Pronunciation[1]

In considering the problems of facilitating inter-lingual communication by language training, it is necessary to break down the task into its major components. This is especially necessary in the case of preparations for programmed learning. In a language communication system there are three major factors involved: (a) a set of signals in the form of visual and/or sound symbols (phonemes, graphemes); (b) a set of combined signals constituting morphological units in the system (i.e., morphemes, words, etc.); and, (c) a set of rules by which these morphological units can be combined into full expressions or sentences (syntax, grammar). Terminology used in the discussion of the structural aspects of language, such as that used in the foregoing description, is common to the field of linguistics (Osgood & Sebeok, 1954; Whatmough, 1957).

The basic task in the teaching of a language is the teaching of a set of signals in which the particular form of communication (language) can be both described and carried out. In everyday communication these signals consist of a set of visual symbols (letters, alphabet) and/or a set of vocal symbols (phonemes or sounds). In a word language the two basic signal systems (the alphabet and the sounds of the language) ideally correspond to each other. If there is a vocal symbol in a language with the sound /d/, for example, it is necessary to have a visual symbol corresponding to the sound. This is exhibited, for example, by the letter 'd' in the English alphabet, the letter 'Δ' in the Greek alphabet, and the letter 'Д' in the Russian alphabet. For purposes of communication it is not absolutely necessary to teach both visual and vocal symbols. In actual practice, however, it is advisable to teach both because both are required for a full handling of a language and a fuller development of vocabulary.

The task of the research reported in this paper was to teach both a limited range of visual symbols and a limited range of associated spoken responses. For obvious reasons this could not be readily accomplished by a simple oral-aural training which would involve no written symbols. The remaining possibilities considered were: (a) "Audio" training, i.e., training by a simple visual representation of the visual symbols with aural representation of the corresponding sounds. This condition suggests language training by presenting the student with simple foreign language words accompanied by the correct pronunciation presented through the ears. (b) "Visual" training, i.e., training by pairing the visual symbols with visual representation of the corresponding sound symbols. This condition suggests an entirely visual training procedure to teach pronunciation ability by using the visual material in which the proper pronunciation is indicated by visual symbols associated with sounds of a language native or well known to the student;for example, the visual symbol "a" in "father" can be used to indicate the Spanish sound /a/. (c) "Audio-visual" training, i.e., training as described in (b) above but using, as an additional technique, a parallel aural representation method as described in (a). This condition would involve both forms of prompting and cueing, i.e., both visual and auditory stimuli in combination to indicate the proper foreign language pronunciation. These three training possibilities suggested three experimental groups for research purposes.

---

[1] This initial report on this experiment was drafted by Attila P. Csanyi, who constructed the programs described in it and who served as chief experimenter on the project.

Certain practical advantages or disadvantages of each of these techniques must not be overlooked. The use of a solely visual form of instruction in pronunciation, if feasible, would be more economical and more convenient than the use of the audio or the audio-visual form, since the visual form requires only the use of a (programmed) textbook. On the other hand, any form of audio instruction would require the operation of some recording apparatus.

The primary aim of the current research was the development of a technique for the visual programming and the investigation of its feasibility. The specific research problem consisted of evaluating the possible differences resulting from the use of the same structural program with its instructional format varied in the above mentioned three ways. On the basis of widely held opinion and common practice it was expected that the audio or audio-visual groups would show the best learning of foreign pronunciation, but it was hypothesized that by the use of a well-structured visual program, the subjects' terminal behavior would closely approximate the behavior of subjects trained with auditory stimuli.

Procedure

Programming Procedure. A programmed text was developed to teach the following -- (a) the Spanish phonemes: /a/, /e/, /i/, /o/, /u/, /r/, /n/, /x/, /k/, /y/, corresponding to the visual symbols: a, e, i, o, u, r, n, j, and the combined symbols qu and ll, and (b) the rules governing the position of stress in

Spanish words. The procedure of programming was inductive, that is, the first visual symbol was introduced with instruction concerning the corresponding vocal response, and the student was asked to read a series of these symbols out loud. The term "vocal symbol" is here used to refer the sound that the student learned to speak while the term "visual symbol" is used to refer to the written material in the foreign language (discriminative stimuli) to which he responds with the appropriate pronunciation. After the student had learned the pronunciation of individual symbols these were combined with others into syllables, and the syllables were combined into polysyllables with instruction concerning the position of stress (accent).

Three experimental methods of instruction were used to define a vocal response corresponding to a visual stimulus. In the visual program, advantage was taken of the fact that a set of corresponding visual and vocal symbols, i.e., a set of English letters and English sounds was already present in the students' repertoire. The method of training in this experimental group involved a transfer of English sounds to Spanish letters with some additional instruction on tongue position in the case of uncommon English sounds or when the English sound did not sufficiently approximate the necessary Spanish sound. Such instruction for the /r/ sound is shown in a sample page of the program in Figure 7.1.

Figure 7.1

Tongue Position Instructions for the Spanish /r/.

A new sound which is very often found at the end of Spanish words is r. To pronounce a Spanish r, you must forget about how you make an r in English and do the following:

1. Say the English word kitty out loud. Kitty, kitty.
2. Now say it again and watch carefully where your tongue is when you come to the "tt" in kitty. Say kitty, kitty, kitty.
3. Place your tongue in the same position where it was when you said "tt" in kitty. Do it ...... say kitty. You find that the tip of your tongue is touching the top of your mouth just behind your upper front teeth.
4. Keeping your tongue in this position, try to make a rolling sound by making the air vibrate through between your tongue and the top of your mouth. It will sound somewhat like a motorcycle. Try it.

Figure 7.1 (continued)

It will be an unusual, foreign sounding *r*. This is the Spanish *r*. In Spanish the *r* is much harder than it is in English. As you notice, it is a *rolling* sound.

5. Go through once more the above steps and try making the Spanish *r* sound several times.

r r r r r r r r r r r r r r r r r r

Figures 7.2, 7.3, 7.4, and 7.5 show four stages of the visual program as used in the visual training group. The program for all subjects was printed on 5" x 8" cards. Examples of these cards are reproduced in the figures mentioned. Figure 7.2 is Page 1 of the visual program and illustrates the first stage in learning to pronounce Spanish vowels a, e, i, o, u. Here a previously learned English word is employed to prompt the pronunciation of a particular foreign language letter. The student is required to look at the English words which are written vertically and to transfer the sound of the vowels written in parentheses to the letter between the horizontal lines. Thus, the student reads "father," "break," "chief," "old," "blue," and approximates the pronunciation of the vowels as in Spanish /a/, /e/, /i/, /o/, /u/.

Figure 7.3 shows Page 8 of the visual program and illustrates how two sounds were combined to produce a syllable. Figure 7.4 shows Page 19 and illustrates the combining of two syllables as well as the teaching of the accent of words. Figure 7.5 shows Page 73 of the visual program which illustrates the training of the subject in the expected terminal behavior of reading actual Spanish words. Notice that throughout the program visual symbols from the student's native language were used to indicate the proper pronunciation of the Spanish sounds. The proper stress of words was first indicated by off-setting the stressed symbol above the line of type (see Figure 7.4 and 7.5) and, as may be seen from the illustrations, this method of indicating the stress as well as the pronunciation cues was withdrawn so that the student might continue to respond appropriately in the presence of the word as it would appear in actual printing.

This visual programming method can be called a "response-transfer" procedure for, in it, a response which heretofore corresponded to a particular symbol(s) in English becomes transferred to another visual symbol. This is the

Figure 7.2
The Introduction of Spanish Vowels
(A Sample Program Page)

Read out loud:

1.
f
(a)
a   a   a   a
t
h
e
r

2.
b
r
(ea)
e   e   e   e
k

3.
c
h
(ie)
i   i   i   i
f

4.
(o)
o   o   o   o
l
d

5.
b
l
(ue)
u   u   u   u

a,   e,   i,   o,   u

Figure 7.3
Combining Spanish Vowels into Syllables
(A Sample Program Page)

Read out loud:

1\.
f
(a)
n a   na   na   na
t
h
e
r

2\.
b
r
(ea)
n e   ne   ne   ne
k

3\.
c
h.
(ie)
n i   ni   ni   ni
f

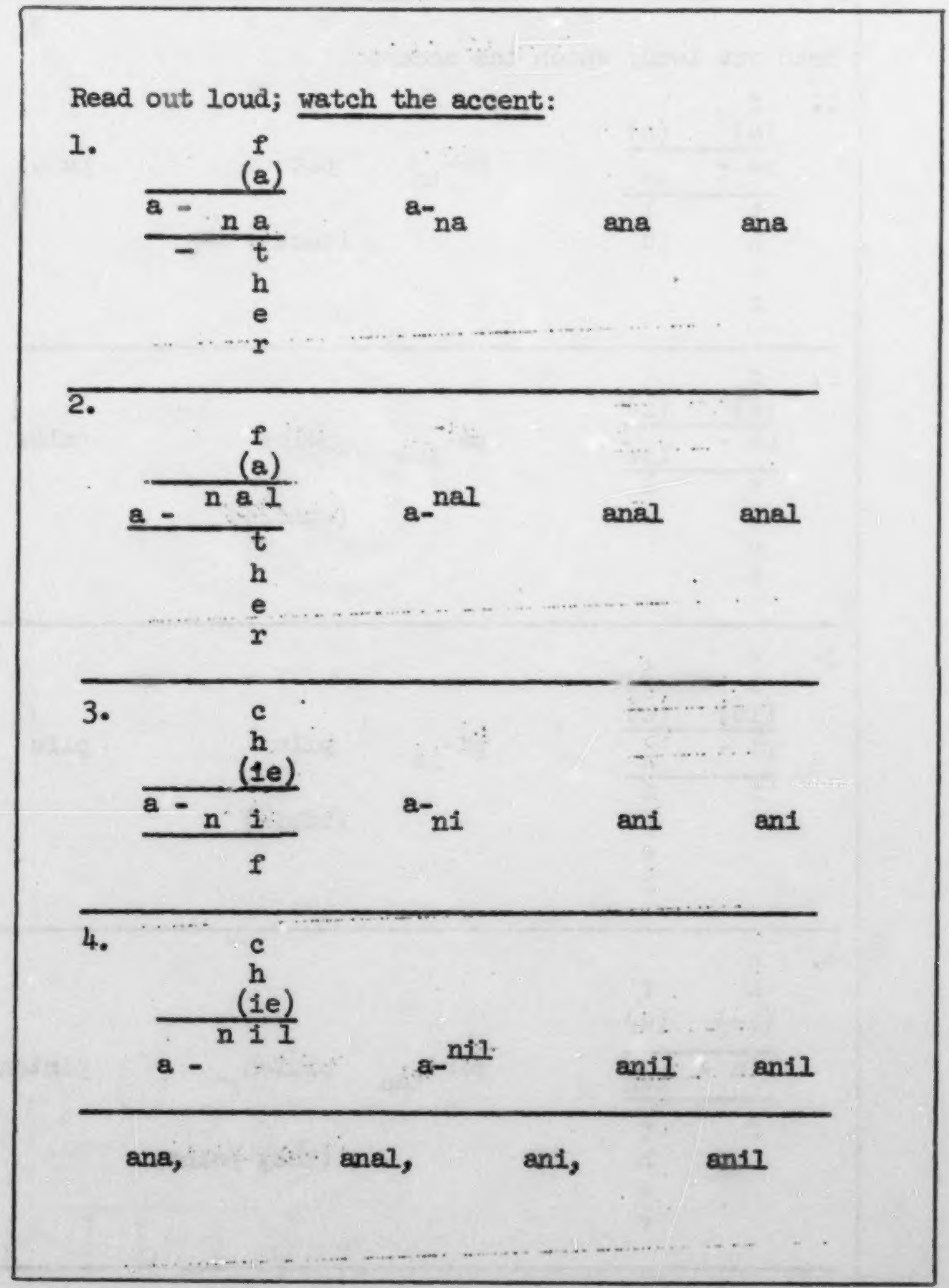
4\.
(o)
n o   no   no   no
l
d

5\.
b
l
(ue)
n u   nu   nu   nu

na,   ne,   ni,   no,   nu

Figure 7.4

Combining Two Syllables With One of the Syllables Accented.
(A Sample Program Page)



Read out loud; watch the accent:

1.
f
(a)
a - n a
- t
h
e
r

a- na   ana   ana

2.
f
(a)
n a l
a -
t
h
e
r

a-nal   anal   anal

3.
c
h
(ie)
a - n i
f

a- ni   ani   ani

4.
c
h
(ie)
n i l
a -

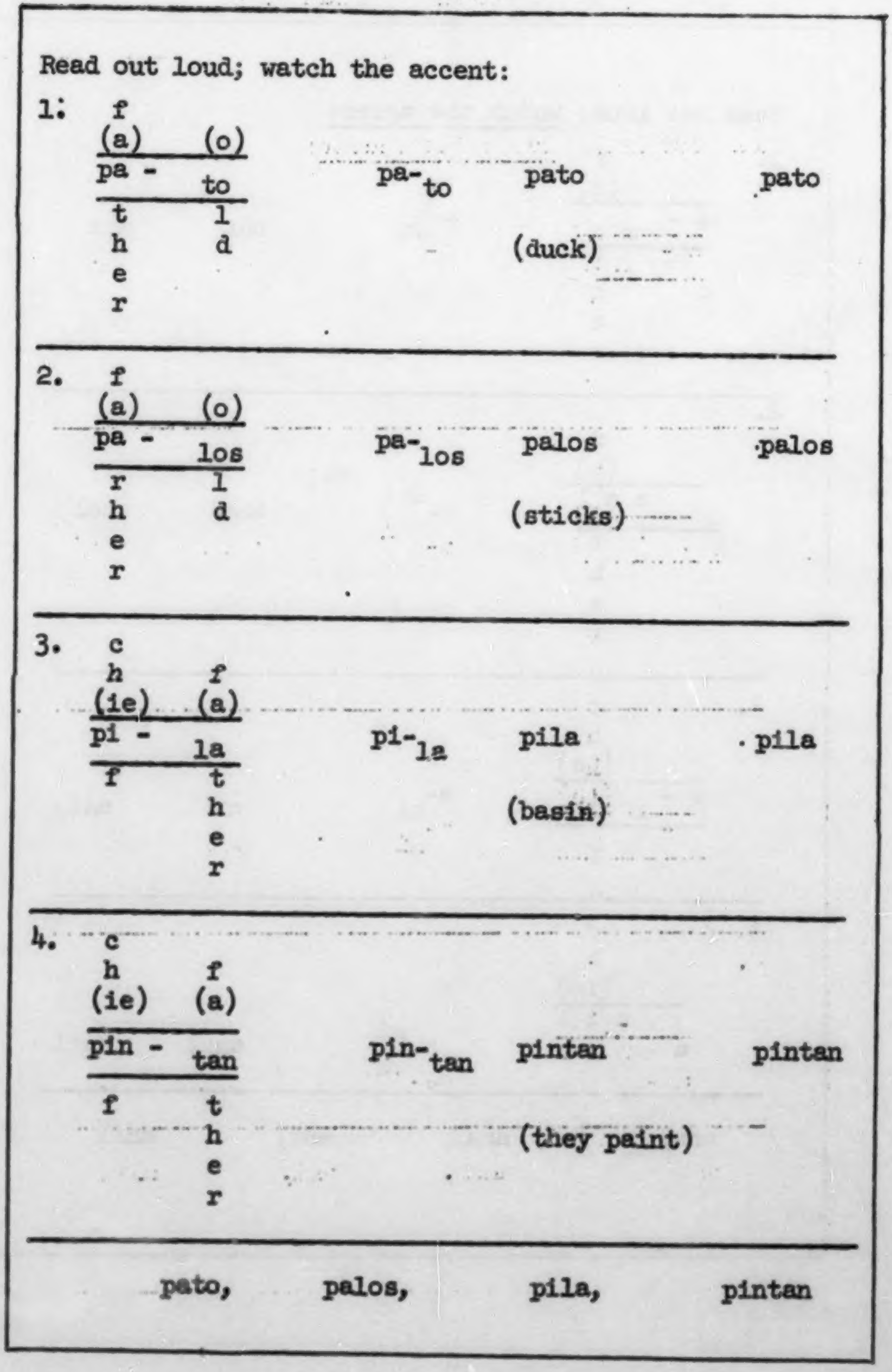
a-nil   anil   anil

ana,   anal,   ani,   anil

Figure 7.5

Reading Spanish Words

(A Sample Program Page)



Read out loud; watch the accent:

1. f
(a) (o)
pa - to
t l
h d
e
r

pa-to pato pato

(duck)

2. f
(a) (o)
pa - los
r l
h d
e
r

pa-los palos palos

(sticks)

3. c
h f
(ie) (a)
pi - la
f t
h
e
r

pi-la pila pila

(basin)

4. c
h f
(ie) (a)
pin - tan
f t
h
e
r

pin-tan pintan pintan

(they paint)

pato, palos, pila, pintan

same response is transferred to a new stimulus. It is also possible to speak of the same phenomena as "stimulus transfer," if one thinks in terms of the stimulus control of behavior. If the term "stimulus-transfer" is used, it implies that the control of the response is transferred from one stimulus to another.

The audio program (the audio group) used a tape recorder for presentation of vocal symbols (sounds) which were to be learned as corresponding to visual symbols. The sequence of the audio program was identical to the visual program described above, the difference being that instead of the response-transfer method, which was entirely visual, the audio-oral method was used. In the audio program the student was presented first with a visual symbol (letter) of the Spanish alphabet on cards and then, through a tape, was told the sound corresponding to this letter and asked to repeat the sound out loud four times while reading the letter being presented.

The third program used in the present experiment was an audio-visual program. This was a combination of the first two programs. It used the same printed text that was prepared for the visual group (using the response transfer technique) in conjunction with the audio tape which was prepared for the audio group.

Subjects. The subjects in this experiment were 30 junior high school students who were selected from a nearby high school population.[2] The subjects were selected on the basis of interest (all subjects were volunteers). These subjects had an I.Q. range of 94 to 124 and had just completed the eighth grade. The subjects were selected from their classes at the end of the regular school year and the experiment itself was conducted during the first week after the completion of the regular term. Subjects were not paid for their participation in the experiment. They were, however, induced to participate in the experiment by being offered an opportunity to learn the use of a tape recorder and to learn some of the things which would be helpful in the first year of regular Spanish classes. (Subjects were drawn only from those students who had registered for Spanish classes for the following year). Subjects in this experiment consisted of five females and 25 males with the females being divided as evenly as possible between the three experimental groups.

---

[2] The authors wish to thank the officers of the Baldwin-Whitehall School District, Pittsburgh, Pennsylvania for their cooperation in providing facilities and subjects for use in this experiment. Especially the authors are grateful for the cooperation of Miss Evelyn Baxter, who, as the person in charge of the language laboratory, greatly contributed to the organization and conduct of this study.

The selection of subjects was carried out by the teachers in the junior high school who provided the experimenters with a list of eligible students. Students had been requested who had no previous training in Spanish and who had no obvious physical or personality disturbances. Thirty subject cards were drawn at random from the pool of available subjects and all 30 selected subjects were ranked on the basis of the most recent Otis I.Q. score. The 30 subjects were then matched on the basis of I.Q. into three experimental groups, assigning the highest score to Group 1 and the next highest to Group 2 and the third highest to Group 3, etc.

Experimental Procedure. The present experiment was conducted in a high school language laboratory. During the running of the experiment the subjects were seated in individual semi-soundproof booths; each booth was equipped with a microphone and a tape recorder. The subjects were seated so that an empty booth was left between each subject, only ten subjects being run at a time. The performance of the students could be monitored on a control panel at the front of the room; that is, the experimenters could select any station by pushing an appropriate switch and could listen to the performance of the student at that station. The students' microphones and the associated monitoring systems were used to assure explicit, overt responding.

All ten subjects in any one of the three experimental groups were run in the same session. Experimental sessions were conducted in two, 30 minute sessions daily for each group with an approximately 25 minute rest break between the two sessions. The running of the experimental groups was staggered in the following fashion: Group 1 began the first day of the experiment at 8:30 a.m. and worked until 9:00 a.m. The subjects in Group 1 would then recess and the subjects in Group 2 began the experiment and worked from 9:05 a.m. to 9:35 a.m. While Group 2 subjects were taking their break, Group 1 subjects returned to the room and completed their second session. Similarly the performance of the third group was staggered with that of the second group.

At the beginning of their first session the subjects were given verbal instructions concerning the procedure for working with the programmed textbook, the microphone, and the tape recorder in the cases of the two groups which were using it. The visual group, which depended entirely upon the visual material for their pronunciation, had no use for the tape recorder and were not instructed in its use at the beginning of the experiment. The students in this group were, however, permitted to learn the use of the machine at the end of the experiment. All further instructions in the experiment were given solely through the programs themselves, i.e., through the written material for the visual group, written material plus individually controlled tape recorders in the case of the audio and the audio-visual groups.

Each subject, working for two 30-minute sessions every day, was permitted to go through the program at his own pace. This meant that subjects finished the experiment at various times. The visual and audio-visual groups finished in an average of 6.1 sessions, the audio group finished in an average of 6.4 30-minute sessions.

Testing Procedure. As each student finished the program he was tested in a separate room, where his responses to a list of 37 printed test words were recorded (see Figure 7.6). In the test-word list the sounds the students had learned occurred in Spanish words of two or more syllables, the sounds occurring in different contexts such as stressed and unstressed syllables, initial and final positions, where it was feasible to do so. Figure 7.6 shows the list of test words as they were shown to the student. During testing each student read the entire list three times. The judgments of the students' pronunciation ability described below were based on the final or third pronunciation of the list. Three repetitions of the list were required to accusstom the student to the use of the tape recorder in the special testing room and to familiarize him with the words, etc.

Figure 7.6

Test Words

Read out loud the following words on this page:

ramo,
higo,
harén,
uñir,
ánsar,
salud,
tajar,
roja,

ñame,
tumor,
amor,
porqué,
étimon,
ñoñería
mohín,
moho

lleno,
reloj,
emir,
aquí,
ídem,
frufrú,
moña,

quema,
llenar,
imán,
tocó,
útil,
puñal,
pollo

jugo,
Japón,
olor,
será,
ómnibus,
hallar,
dique,

Evaluation and Judging. The recordings of each student's pronunciation of the test items were evaluated by four independent judges.[3] Two of these judges had been teaching Spanish at the high school level, and a third at the college level. The fourth judge, as well as the college teacher, were native speakers of the Spanish language. In the course of 3 to 4 sessions, each of the judges listened to and evaluated the performance of every student on every word of the test list. Repeated judging sessions were made necessary by the fact that the judges spent a total of approximately seven hours apiece in judging the students' performance. Each subject was assigned a code number under which his performance was evaluated. On those occasions where more than one judge worked at a time judges were seated so as to lessen the probability of their influencing each other's judgments either through spoken or visual communication.

---

[3] The authors are grateful to the following individuals who served as judges in this experiment: Miss Evelyn Baxter, Language Coordinator for the Baldwin-Whitehall Public Schools and Director of the Language Laboratory; Dr. Matilda Contreras, Professor of Spanish, University of Pittsburgh; Miss Patricia Previtt, teacher of Spanish, Wilkinsburgh High School; and Mr. Pedro Trullas, a student at the University of Pittsburgh who is a native of Spain and a resident of Venezuela; as well as to Miss Donna Gow for her aid as a research assistant in running the experiment, conducting the rating sessions with the judges, and analyzing the data.

The students' spoken words were played one at a time so that the judges could hear and evaluate each word. Evaluations of each word were made on a rating scale of 0 to 4, from 'incorrect' to 'like native' pronunciation. Separate judgments for each word were made of: (a) each individual sound; (b) the position of stress; and (c) the 'overall' quality of the word. The five categories of judgment for each sound and word were as follows: 0, incorrect, 1, passable, 2, understandable, 3, good approximation to native pronunciation; and 4, like native pronunciation (see Figure 7.7). Judging proceeded as follows: Each judge was given a booklet in which each page contained rating scales for two words. The first word was played, and the judge made his evaluation of the sounds, accent, and overall pronunciation by placing checks along appropriate scales under the categories mentioned above. When the first word had been judged, the second word was played on the tape recorder and the second word was rated. Each judge completed a separate book for each of the 30 students. An example of a page from the rating booklet is shown in Figure 7.7). This is the first page of the rating booklet and shows the rating scales which the judge was to fill in for the words 'ramo' and 'ñame.'

At the beginning of the first judging session with all judges present, the judges were instructed in the general judging procedure and in the meaning of the categories. They were instructed not to communicate with each other regarding the evaluation of the students' pronunciation. For the most part, following the first rating session, judges came for subsequent rating sessions individually at times of their own convenience.

Summarizing the measurements available in the present experiment: four judges rated 37 words repeated once by each of 30 students. The words were rated in terms of separate sounds, accent placement, and overall word quality. This yielded a total of 194 evaluations per subject per judge. As indicated, the dependent variables--measured in the present experiment were sounds, accent, and overall quality of pronunciation. The present report considers only an analysis of the mean score obtained by each subject from the ratings of all judges on all dependent variables.

Figure 7.7

Sample Page from the Judges' Rating Booklet

ramo

	In-correct pron.	Pass-able pron.	Understand-able pron.	Good approx. to native pron.	Like native pron.
r	0	1	2	3	4
a	0	1	2	3	4
o	0	1	2	3	4
accent	0	1	2	3	4
overall	0	1	2	3	4

ñame

	In-correct pron.	Pass-able pron.	Understand-able pron.	Good approx. to native pron.	Like native pron.
ñ	0	1	2	3	4
a	0	1	2	3	4
e	0	1	2	3	4
accent	0	1	2	3	4
overall	0	1	2	3	4

For each student, then, a single value was obtained which could fall between the limits of 0 and 4. This total performance score was the mean of 4 x 194 or 776 scores given for each subject by the four judges.

As a check on the agreement between judges, correlation coefficients were computed between judges' ratings using the 'overall' pronunciation rating of the words in the test list. These correlations ranged from .61 to .85 indicating an acceptable level of inter-judge reliability.

The obtained mean values for each student were entered into a simple analysis of variance, the results of which are shown in Figure 7.8. In this table, the obtained F of 1.4 does not reach the .05 level of significance, thus the hypothesis that there are no significant differences between experimental treatments cannot be rejected. The mean scores for the three groups were: visual, 1.56; audio-visual, 1.36; audio, 1.22; indicating for all groups received ratings between "passable" pronunciation and "understandable" pronunciation. The visual group, the group for whom correct pronunciation was indicated by purely visual symbols, was slightly superior to the other two groups. As seen from the analysis of variance, however, this was not a statistically significant difference.

Figure 7.8

Analysis of Variance Between Experimental Methods

Source	df	Variance Estimate	F
Between methods	2	0.28	1.4
Within methods	27	1.20	
Total	29		

Required for df $n_1 = 2$, $n_2 = 27$: F = 3.35 for .05 level.
F = 5.29 for .01 level.

Although the present report does not deal with the analysis of specific measures such as particular sounds and accent, an inspection of these specific ratings showed that the visual group was rated superior to both of the other groups on measures of all the variables but one, which was the letter 'ñ' where the audio-visual group scored higher. The audio-visual group was superior to the audio group on measures of all the variables but three, namely the letters 'r' and 'j', and overall quality. Complete and detailed evaluation of more specific aspects of the data awaits further analysis.

The above results indicate that an oral response repertoire can be developed through visual prompting techniques by using the 'response-transfer' procedure. The study suggests, that visual programming can be employed to accomplish some of the end results that can be obtained by audio programming. Whereas the program used in this study does not claim to teach native pronunciation to the student, it does provide approximations to native pronunciation from which the teacher can shape the student's behavior towards perfection. The question arises of how the teacher can accomplish this aim, and further research is needed to develop methods and techniques which might result in specific recommendations to be suggested to language teachers. In terms of programming techniques further analysis of the present data should reveal weaknesses of the visual prompting program which can be corrected and taken account of in a revised form of the program.

References

Whatmough, J. Language: A modern synthesis. New York: The New American Library of World Literature, 1957.

Osgood, & Sebeok, T. A. Psycholinguistics: A survey of theory and research problems. Supplement of Int. J. of Am. Linguistics, 20, 4, 1954.

Chapter 8

A Vocabulary Program Using "Language Redundancy"[1]

Learning a new language involves the acquisition of a variety of new skills which may be separately identified: pronunciation, grammar, vocabulary, elegance of use, to name but a few. Recent attempts have been made to teach the pronunciation of a foreign language without regard to grammar or meaning of the new sounds that are being uttered (Morton, 1960). These attempts have been quite successful. No one, however, has undertaken to teach the vocabulary of a language without regard to pronunciation. The technique presented in this chapter lends itself to teaching a vocabulary and also a knowledge of the new grammar of a foreign language. While it is probably possible to teach vocabulary without any recourse to the new grammar, there is little doubt that to teach vocabulary without some reference to grammar is more difficult and of less value to the student.

The technique described does not replace conventional language learning. Programs prepared according to this technique probably will serve to enrich existing language courses in the sense that they allow the student to use a large variety of words from the very beginning in a meaningful way. The terminal behavior which can be achieved by this method is the understanding of sentences in the new language in a written form. A simple newspaper story, for example, can be read by a student who has gone through a program prepared in accordance with the technique to be presented here.

Basic Concepts Underlying This Form of Programming

It is rare in the every day situation that a control stimulus is not composed of many elements. Even so simple a stimulus as the green light which controls walking behavior at an intersection is inter-connected with so many other properties of the environment that the seasoned pedestrian usually attends to the accompanying stimuli more than he does to the green light as such: he will begin to walk, for example, when the amber appears on the other direction. He "knows," one might say, that green is due to appear now at any moment. In terms of information theory notions there is enough redundancy in the traffic light situation to communicate even in the presence of considerable noise.

---

[1] The original paper was prepared by Halmuth Schaefer and was issued as a separate project report by the University of Pittsburgh, January 1961.

Language is a prime instance in which redundancy is used to permit communication or control over behavior in the presence of noise. Many fewer sounds than those conventionally used in grammatically and phonetically correct speech are necessary to communicate. For example, the sentence "A line may be either curvilinear or straight" does not really need the word "or" for anything but a filler. The word "or" gives no additional information. It could be (and in certain codes it is) simply omitted.

In behavioral terms the words "either...or" constitute a control stimulus which need not be presented completely to be effective. A response to the sentence "A line may be either curvilinear ______ straight" would be essentially the same as that to "A line may be either curvilinear or straight." Rather than omit the word "or" in the above sentence it is replaced by the "nonsense" form which happens to be the equivalent word in a foreign language we wish to teach, for example, "A line may be curvilinear oder straight." "Oder" can have but one meaning in this context. Since it is understood that the subject is learning a new language, a justified conclusion would be that "oder" is another word that means "or."

Often the situation is not so simple. It may be difficult to find a sentence in which only one interpretation of a new "nonsense" stimulus is possible. In these instances it is necessary to rely on subsequent recurrence of that stimulus in other connections which will not allow the wrong translation or substitution to survive. This is the process of stimulus discrimination which is an effective method of learning. Take for example: "A Kind follows the normal process of growth." It is possible here to learn that Kind means plant, tree, child, bacteria, city, or any other number of things. There is no telling which of these alternatives would be learned initially. The probability that the wrong one be learned is rather great. Subsequently, however, a sentence may occur that reads: "A Kind whose mother and father were dead..." this excludes some of the meanings possible in the previous sentence such as plant, city, tree, etc. In this way, the control function of the stimulus and the effect that it has over the speaker can be limited. But it is still possible to substitute cat, dog, boy, girl, etc. Other sentences, however, will occur containing the word "Kind" which gradually will extinguish all but the correct response "child" to this stimulus. The last responses to disappear in this case very likely are the responses "boy" or "girl" to the stimulus "Kind."

It might be argued that this is a very tedious and time consuming way to learn a simple word. It must be remembered, however, that words rarely are simple. The word in the above example is of neuter gender. It is a literal translation of the English "child," however it is used in the German language in certain situations where the word "child" is not necessarily used. The opposite is true also. All of these things about the new word would have to be communicated in some other fashion if the student were to learn how to use the new word. To obtain this kind of knowledge, much of it quite unformulated, the method of "letting the student find out for himself" seems much more appropriate than to describe in lengthy instructions, if such could be given, the usage of the new word.

Behavioral analysis must ask what the reinforcement is that makes learning under these conditions possible. The most obvious reinforcer in this situation might well be the contents of the reading matter which captures the reader to such a degree that he is willing to make the slight effort of tying the new word in with the known sentence contents. Another reinforcer, sometimes important and sometimes of no effect, is the hope to be able to read another language. The beginning of a program based on the redundancy method is almost 100 percent in the native language; but the end of such a program is 100 percent in the foreign language. This quality of the program is pointed out to the student in some introductory statement. Lastly, there is a reinforcement of the general tact[2] of "getting something for nothing," or "getting by with something." This last kind of reinforcement probably works because learning a new language is traditionally beset with aversive stimuli of many kinds. It is very time consuming and the student is more often punished for being wrong than reinforced for being right. Any adult embarking upon the task of learning a new language knows this. With the present technique of programming, however, the terminal behavior (reading in a foreign language) might be achieved in a relatively short time--measured in weeks rather than in months, or years--and there is no punishment for being wrong. (To human beings, exposed for the most part to aversive control, this may seem like a subtle form of cheating. They may express fears that "this can't work," but upon seeing that they have nothing to lose are willing to possibly "waste" a few hours. They feel that, if contrary to their better knowledge, this method proves to be successful with them, they have, indeed, gotten "something for nothing," an event which is to many people highly reinforcing.)

---

[2] The theoretical basis on which this programming technique was developed is Skinner's Verbal Behavior (1957). A tact can be defined as a verbal operant in which a response of a given form is evoked (or at least strengthened) by a particular object or event, or property of an object or event.

The Techniques of Introducing the New Words of a Language

One may distinguish between three different kinds of words which are dealt with in any language: (a) skeleton words (a, the, and, but, or, in, out, have, take, make, etc.); (b) common words (table, chair, floor, boy, girl, etc.); and (c) rare words (extrapolate, manger, joist, mandipular, etc.). There is no general rule that can be given as to which words of these groups should be introduced first. Generally speaking, however, it is not unwise to begin with the skeleton words. These occur so often that the programmer hardly has to worry about the context in which he introduces them. Any lack of discrimination that may exist early due to poor programming is certain to disappear as the text progresses.

The most difficult words for the programmer are not the rare words, as one might suppose. These words usually occur in a context which makes it very easy to rely on the control exerted by the context of the sentence in which they are likely to occur. This is called a thematic prompt in programming. The word "manger," rare as it is, can easily be identified in our culture in a sentence such as "Christ was born of a virgin in a Krippe." Idioms in the known language will serve excellently to introduce rare words because they form part of the existing response repertoire. The idioms of the new language, of course, have little or none of this control. Often they sound farfetched. English, for example, has "to put the cart before the horse." The English idiom in this case lends itself during any stage in the programming to get at such words as "horse," "before," and "cart."

This leaves the common word group, and it is evident that greatest care and ingenuity on the part of the programmer are called for to find several stimulus situations for each of these words in which but one interpretation is possible. In fact, in introducing the common words it is most evident that this form of programming requires both linguistic knowledge and versatility in behavioral analysis. There are more ways than one in which the correct word can be "guessed" in a sentence. In introducing common words the programmer must rely more than in any other case on the process of gradual stimulus discrimination. He is free, for example, to introduce additional sentences in a text which repeat what has been already communicated in a slightly altered form and which tend to extinguish wrong interpretations of a newly introduced word. He must also use his judgment to decide how important it is at this particular point in the program that the subject know the word very precisely. Sometimes this is not crucial. There are situations where it makes little difference in the understanding of a story whether a particular word is read as "child," "boy," or "girl." There are, however, other situations where it is definitely important that a word be understood in its specific meaning since the rest of the story does not make sense any other way or would be distorted if the wrong meaning were used.

The program itself can be seen to fall into three distinct phases. In the first phase the programmer will introduce many of the skeleton words. In doing this he can do no more than plug them in at the appropriate places. The most efficient way of proceeding at this stage of programming is for him to cross out all those words in the English translation of the text--or the English original that is being used--that he wishes to replace. The typist then types from the printed page leaving underlined blanks at all those places where a word has been crossed out. As the programmer receives the typescript he fills in the blanks and enters all the words (in the new language) in a ledger with dividers from A to Z. Every time he inserts a word he notes in the ledger how often the word has occurred. Proceeding in this manner permits a check on whether the sentence is strong enough to make clear the meaning of a particular word that is to be inserted. This is often not obvious on first analysis, i.e., when the programmer crosses out words in the original text. If, as he receives the typescript with the blanks, he finds that he has no difficulty in substituting the correct words, his original analysis has been satisfactory, and, for the most part, there should be little difficulty on the part of the student in "guessing" the correct word. The typescript with filled in blanks is then handed back to the typist who prepares the final program with blanks filled in with words of the new language. An example of the first phase is:

> True! - nervous, very, very dreadfully nervous, ich, had been, and am; but why will you say that ich am mad? The disease had sharpened meine senses - not destroyed, not dulled them. Above all was der sense of hearing acute. Ich heard all the things in dem heaven and der earth. I heard many things in hell. How, then, am ich mad? Hearken! Und ebserve how healthily--how calmly ich can tell die whole story.[3]

---

[3] Edgar Allen Poe, "The Tell-Tale Heart."

The *second phase* of the program is that in which many of the skeleton words are assumed to be known by the reader already. Several common words have been introduced also. It is in this stage that the programmer begins to couch sentences in the grammatical form of the new language. Very few new words are introduced in this phase. Instead the subject learns to understand sentences when they appear in a grammatical order with which he is not familiar. There will, of course, be occasions in the writing of this second phase to introduce new words. These occasions should be made use of as they occur.

It is not possible to give a hard and fast rule as to when the second phase begins, nor to say how long this phase should last. Naturally, it depends on the skill of the programmer on the one hand to cleverly shape single sentences using the existing response repertoire established up to this point, and, of course, also upon the special features of the language to be used. An example of the second phase is:

> *Der* second *und* third day went by *und* yet showed himself *mein* tormentor *nicht*. Again could *ich* as free man breathe. *Das* monster was apparently in great terror run *away*! Never again would *ich es* see! *Meine* happiness was complete! *Die* guilt *der* black deed *disturbed mich* but little. Some questions were *asked und* readily answered. *Eine* search was even undertaken, but, of[4] course, could *nichts* be found. *Ich* looked *einer* safe future toward.

Sentences which violate all feelings of harmony and propriety, which most lovers of language have, will thus be frequent in the second phase. The programmer, nevertheless, must make use of grammatical habits of the subject for the sake of reinforcing the reader whenever he introduces the words. It is well for the programmer to remember that in using either the words of the new language in an impossible form, that it is the feelings of the accomplished speaker, not those of the learner which he offends. He should remember that the program is not written to please the programmer himself or the expert critic. It is written solely to teach the student a new response repertoire. This new response repertoire consists of being able to read the understand the new language. To obtain this behavior it is necessary to go through stages which will definitely be "wrong both from the point of view of the speaker of the known language and from the point of view of the native speaker of the new language. But the programmer known that he must gradually shape the new behavior that he wishes to establish. Unlike Athena who sprang full-grown from the head of Zeus, the product emerging from a program goes through stages of making. These stages are quite visible in a program (they are not so visible in a conventional language teaching) and they look distorted. Conventional language teaching often strives to have the student use the new language in a perfect form from the very beginning. This results in the introduction of childishly simple sentences of the type, "La plume de ma tante est sur le bureau du mon oncle; la papier de mon oncle est sur le bureau de ma tante." It is of interest to observe, however, that while we call such sentences childish they are hardly the kind of sentences that a child would utter. On the other hand, the kinds of sentences that the student will form in the present technique of programming may, especially at this stage, sound very much like the bablings of a child. The student is introduced to the new language in a way not unlike the way a child learns to speak. To say it in a general form: it is the terminal behavior that the programmer must have in mind; how he achieves it is of less concern; if he has to teach, temporarily, "wrong" habits in order to get there, he is perfectly willing to do so since no harm appears to be done. He will not offend or hurt the student learner. He will, quite likely, offend his own sense of aesthetics and that of the accomplished speaker of either the existing or the new language. This, however, is of less concern than accomplishing the task.

---

[4] Edgar Allen Poe, "The Black Cat."

In writing for the second phase of the program it is most convenient for the programmer to utilize a dictating machine. He may either use the words of the new language as he dictates, or he may simply dictate "blanks" if his typist is not fluent in the new language. There is the possibility of spelling the new words of a foreign language so that even an unskilled typist could perform; however, it has been found that the burden placed upon the programmer by requiring this additional work is disturbing to the point that dictation becomes very cumbersome and confusing. It is probably better to dictate with "blanks" and then edit the typescript by inserting the proper words.

For the third phase the programmer has now available a small repertoire of words in the new language which he must use skillfully to form sentences in the proper grammatical order of the new language. This is at once the most rewarding and also the most difficult stage. Sentences have to be rewritten and simplified for the sake of the existing response repertoire which is contained in the words appearing in the ledger. The ledger must be consulted continuously. New words are introduced, as in Phase 2, if the content of the sentence makes this feasible.

But the third phase will also contain many sentences in the proper grammatical order of the native language for the sake of introducing new common words. Gradually, as vocabulary increases, the text will be more and more in the new language. At some time during the third phase, and perhaps during the second phase, the programmer does well to use an existing translation of the literature that he utilizes for his program. An example of the third phase is:

> Die slope seiner Wande wurde von Moment zu Moment smaller, und der bottom der Vortex seemed sich gradually zu lift. Der sky war klar, die Winde hatten sich died, und der moon went brightly im Westen down, als ich mich auf dem surface des Ozeans facing die coast von Lofoden found, exactly über der place, wo der Trichter des Moskoestromes gewesen war.[5]

There are general considerations which are almost obvious: the programmer should keep track of the number of times that he used the word of the new language in his program only up to a certain point. As a first approximation, initial tryout suggested that if he notes that a word has been used 10 times it is not necessary to indicate any further the times when the word is used, since it can be assumed that the word is known after it has been used in context 10 times. (As indicated later, subsequent data showed little consistent relationship between number of repetitions and learning.) Since it may be of some interest to know approximately during which part of the program a particular word was used and since the program typically consists of a series of short stories, it is wise to indicate in the ledger at what section of the program which particular words have been introduced and practiced. The following is an example of a ledger:

WORD	STORY 1	STORY 2	STORY 3
Katze	1,2,3,4,5,6		
Kater	1,2,3,4		
klar	0	1,2,3,4	5
Kreatur	1,2		
Keller	1,2,3,4,5,6,7		
kein	1,2,3,4	5,6,7,8,9,10	
konnte	1,2,3,4	5,6,7,8,9,10	
kommen	0	1,2	3,4,5
kam	0	1,2	3
gekommen	0	0	0
kann	0	1,2,3,4,5,6,7,8,9,10	
klein	1,2	3,4	
können	1,2,3,4,5,6,7,8,9,10		
Konstellation	1		
Knie	0	1	
Kopf	1,2		

---

[5] Edgar Allen Poe, "The Descent into the Malestrom."

Modes of Presentation

The program appears in book form in either one of the following ways:

Mode A. No overt response required. In this method the reader is simply presented with a text which gradually shifts into the new language. A dictionary may be provided with the text (a dictionary which is no more than a reprint of the words collected in the ledger) which may be used by the student as he is reading. The hypothesis underlying this form of presentation is that the use of the dictionary may be considered a highly aversive behavior which ought to extinguish as other behaviors become available which are less aversive and can be used to achieve the same result. As the reader goes through the text, he should find that using the dictionary slows him down and he will avoid this effort and spend time on guessing what the word means on the basis of the context in which it is used. The risk in this kind of programming is primarily that the alternate response to dictionary using, that of translating from context, is not reinforced early enough to persist and to replace the undesirable response. There are as yet no empirical data to substantiate the effectiveness of this mode of presentation. It is, however, probably a good method for students who have learned other languages and who will appreciate methods of learning new words which do not involve the dictionary.

Mode B. Write in the first 10 times a word occurs. This mode of presentation provides a space at the margin of the text where the student is required to write in the word which is on the corresponding line. The known language term should never be given. An example of this mode of presentation is:

> "Angel Holiday was die kind of woman who abhored ____ = ____
> no means to achieve an end. Following her phi-
> losophy she achieved die following ends: a ____ = ____
> penthouse on der seaside, two outrageously expen- ____ = ____
> sive mink stoles, ..."

Shortcomings of this method are that the student may make up his own diction of the words which may be at odds with the proper pronunciation. Whether this is commonly true or not is not known at the present time.

Mode C. Write in the first 10 times with pronunciation the first time. This means that the foreign word must be given once more at the margin. An example of this mode is:

"Angel Holiday was die kind of woman who abhored die no means to achieve an end. Following her philosophy she had achieved die following ends: a penthouse on der seaside, two outrageously expensive mink stoles, ..."

(DEE) = ______

______ = ______

der(DARE)= ______

There are theoretical objections to this method of proceeding which need not be considered here.[6] Again it must be remembered that the terminal behavior dictates the methods that the programmer employs. In this case the terminal behavior consists of being able to read a newspaper in the new language and this is the the end to be accomplished. Since it is felt that a subject, in reading through the text, will make up his own way of pronuncing words as he reads them, it is not a bad idea to give him hints as to what is an acceptable pronunciation. It is true that he does not learn to pronounce the words as a native speaker would, but it is certainly true that he will reach a stage of approximation which, for the time being, may be regarded as satisfactory. At least, he certainly does not learn radically wrong pronunciation and basis is provided for subsequent refinement of pronunciation if desired.

Preliminary Tryout and Discussion

A program was prepared introducing 350 German words in the context of three stories by Edgar Allen Poe: "The Tell-tale Heart," "The Black Cat," and "The Descent into the Maelstrom." Three subjects were given these stories to read. The mode of responding selected was that in which the first introduction of a word is accompanied with a bracketed phonetic writing at the margin. The subjects were encouraged but not required to write in the words as they appear. All of the subjects, two female and one male, were undergraduate students. They had no previous knowledge of German and no member of their family spoke German.

The subjects' reading time for all of the stories was six hours, five hours, and 4.5 hours respectively. After the subject had finished reading all of the material he was given two tests. For Test 1, the subjects were presented with a random sample of the introduced words out of context, and they were asked to give translations of these words. In all, the list contained a total of 115 German words not counting many variations on the root words such as verb conjugations, etc. Table 8.1summarizes the results of this word recognition test for each of the three subjects. In interpreting the accuracy data in Table 8.1it should be born in mind that not all of the words on the test had appeared in the text material the 10 times; only 24 of the words on the test met this frequency requirement and many of the words had been exposed in context to the students only one or two times. In view of the fact that the three stories used did not permit the completion of full programming for all words, the overall accuracy percentage of 66 seems quite good. The second test consisted of presenting the subjects with simple sentences (see Table 8.2)all of which they translated without difficulty. The success with which the subjects could pronounce the words varied greatly. It should be remembered that no formal programming was undertaken to teach pronunciation. The only reason for including the phonetic writing of words was the observation that subjects make up their own pronunciation as they go along, and in the absence of anything better the approximation on the basis of English spelling was given.

---

[6] Even the best phonetic system does not teach proper diction in a new language since the subject has not yet learned to discriminate the sounds of the new language. To the non-Latin speaker the "ah" sound that he makes as he pronounces the word father is only an approximation to the Spanish "ah." (See Chapter 7).

The subjects who took the program had at no time any feeling that any hardship was imposed upon them. Two of them had read the stories in question previously but they did not recall exactly what they were about. The other subject had not read the stories. This made little difference in the interest which the subject showed for the stories. Yet, at the end of a relatively short time they had a response repertoire which they previously had not had. In addition, however, they also knew quite a bit of grammar in the new language even though they could not formally state grammatical laws. One subject who had read passages of the text out loud had begun to pronounce, in German, English letters such as "a," if they occurred out of context. All of these things, of course, could be studied by experiment.

Table 8.1

Summary of Subject Performance on Word Recognition

	Missed	Wrong	Correct
Subject B.	18	20	80
Subject W.	33	16	72
Subject P.	20	12	83
Total	71	48	235
Total (Percent)	20%	14%	66%

Table 8.2

Sample Test Sentences

Ein alter Mann hatte ein Haus. Ein alter Mann hatte eine Frau. Am 12 Juli 1958 war ich in New York. Es war ein warmer Tag. Ich sah die gigantischen Häuser in New York. Im Ozean sid viele Fische. Ein starker Wind blies über den Ozean. Ein Haus hat vier Wände. In der Wand ist ein Fenster.

Further analysis of the data showed that the number of times a word had been introduced into the text had no bearing on the surity with which it was learned. This is not especially surprising in view of the fact that many of the words bear a close resemblance to the English and that others have special qualities through existing verbal chains. It seems that it is not sheer frequency of repetition which teaches a particular word. Rather, the properties which individual words have acquired in the context of usage are the ones which assure learning.

Any further investigation that is called for in this context would profit from following the pattern of analysis of verbal behavior suggested by B. F. Skinner (1957) in his book Verbal Behavior. It should be emphasized that the present program was written by a person who is not a trained linguist, but the program had shortcomings due to the lack of knowledge in linguistics, and that the science of verbal behavior itself is extremely young. However, the preliminary results are encouraging enough to indicate that this method of programming deserves further attention.

References

Morton, F. R. Guidelines and first progress report on NDEA, Title VI, Project ALLP (SAE #8938). University of Michigan, December 1960.

Skinner, B. F. Verbal Behavior. New York: Appleton-Century-Crofts, 1957.

Chapter 9

An Exploratory Evaluation of a Discriminative Transfer Learning Program Using Literal Prompts [1]

This chapter describes a short learning program that was constructed as an example of what seems to be a general program type. The program is intended as an illustration of one approach to programming in situations where the behavior to be learned involves discrimination training. A learning program or a part of a program is usually designed to carry out specific behavioral processes, e.g., acquisition of a response or the discrimination of a stimulus. Most available programs show a rather difficult to describe combination of such basic behavioral processes. For the purpose of research in programming it seems fruitful to isolate the separate functions of programmed sequences so that they can be subjected to individual study and so that several functions can be most efficiently combined.

Investigations of this kind, i.e., the study of programming techniques relevant to particular learning processes, are required for the translation of behavioral principles into instructional procedures. Program writing will be improved to the extent that the program writer can use a working knowledge of the behavioral principles involved. This paper is an attempt to show how one such basic principle, discrimination training, was applied in a rather "pure" form to the construction of a program.

The application of the discriminative training paradigm to human learning programs rests on simple assumptions and permits a clear specification of the procedure to be followed. In bringing about discriminative control it is simple required that a response be elicited in the presence of the stimulus which is to acquire discriminative properties, and that reinforcement be contingent upon such an appropriate occurrence. It must be assumed at the beginning of discrimination training that the response over which control is sought exists at some strength

---

[1] The authors are indebted to Miss Ruth Spangler who, while she did not participate in the writing of this report, should properly be listed as a co-author of the program itself. Appreciation is also due to Miss Magdalen L. Itzel of the Baxter Elementary School (Pittsburgh, Pennsylvania) and to Mrs. Bertha Seaman of Overlook Elementary School (Baldwin-Whitehall Schools, Pittsburgh, Pennsylvania); these principals generously permitted and encouraged program testing in their schools. This report is substantially the same version as the project interim report, March 1961.

strength in the repertoire of the organism. If the response is weak or improbable it can be elicited in the presence of the discriminative stimulus by presenting, in addition, some stimulus which previously was brought to control the response. Such auxiliary stimuli which are initially used to prompt the response may, as training proceeds, be gradually dropped. The control of the response in this way is transferred to the new stimulus. This is the basic paradigm upon which the program to be reported in this paper was based. With the human organism it is possible to carry on several such training procedures simultaneously. The auxiliary stimuli which are initially depended upon to produce the response upon the appropriate occasion have a variety of forms. In the area of human verbal behavior it is possible to substitute one stimulus for another in most cases where both stimuli evoke the same form of behavior. For example, a picture of a house and the word "house" can both be used to evoke the same verbal response. If one wished to bring the vocal response under the control of the written word, the discrimination training paradigm suggests that the picture be presented together with the word and that the picture be gradually faded or dropped from the situation. This switching or transferring of stimulus control in programmed learning is in line with suggestions made by Skinner (1959) who proposed a situation in which verbal responses could be attached to non-verbal stimuli as in the learning of geography or anatomy.

Although Skinner (1957) has treated the problem of prompting in verbal learning, the actual technology of prompting is only beginning to receive experimental attention (e.g., Angell and Lumsdaine, 1960). It is possible to state several simple procedures by which the prompts may be dropped from a learning sequence. One method of removing the auxiliary stimulus has been called vanishing. In this case, the term vanishing is best taken to mean that the prompt is literally vanished, part by part. Fading is another technique of prompt removal and again, a literal interpretation seems to serve best. In fading, the prompting stimulus would be actually faded in intensity, saturation, or some other physical quality while all other parts of the prompt remain the same. A third, infrequently employed, technique of prompt removal is prompt distortion. Using prompt distortion, for example, it would be possible to elicit the word "camel" with a picture of a camel and gradually distort the picture into the written letters that spell the word. As research on programmed learning proceeds, these and other methods of prompting will be further developed if they are found to facilitate learning.

The question which the program described in this report was designed to answer was whether or not a simple application of the discriminative transfer training paradigm, making use of literal prompts, would provide an effective instrument for use in the area of verbal learning. The programming technique developed involves: (a) the transfer of control from a previously learned and effective stimulus to a new stimulus, called discriminative transfer training, and (b) the actual vanishing of portions of the previously controlling stimulus. This combination is called "discriminative transfer programming using literal-prompt withdrawal."

The Program

The program described here was one called "Learning Color Names" which was designed to teach the reading of eight color words to beginning first grade students.[2] The program capitalized on two simple facts. First, that a color may be represented in two ways, either by the color itself or by the word-name for the color. Both the color and the name have the same verbalization as an appropriate response. Secondly, children learn correct verbalizations to color before they learn correct verbalizations to color names. Observing the useful dictum that a program should use existing behavior as a basis for building new modes of behavior, the task of the program was easily specified. A program was required which would transfer or extend a verbal response from one stimulus (a wave length of light) to another (a word). The program would have to arrange as many stimulus response extensions as there were words to be taught, e.g., one each for red, green, blue, etc. It should be emphasized that this was a program designed to exemplify discrimination training and that programs designed to carry out other functions such as the acquisition of a novel response, stimulus generalization, or response differentiation would probably be quite different in appearance from the one presented here.

The "Learning Color Names" reading program was designed to teach children to read the names of eight common colors: brown, black, purple, blue, orange, green, yellow, and red. The behavioral prerequisite to successful use of the program was the ability to name the colors correctly when a colored object was shown. This program taught "reading" in the sense that the child learned to respond vocally to a specific set of letters in a fashion common to the adult level.

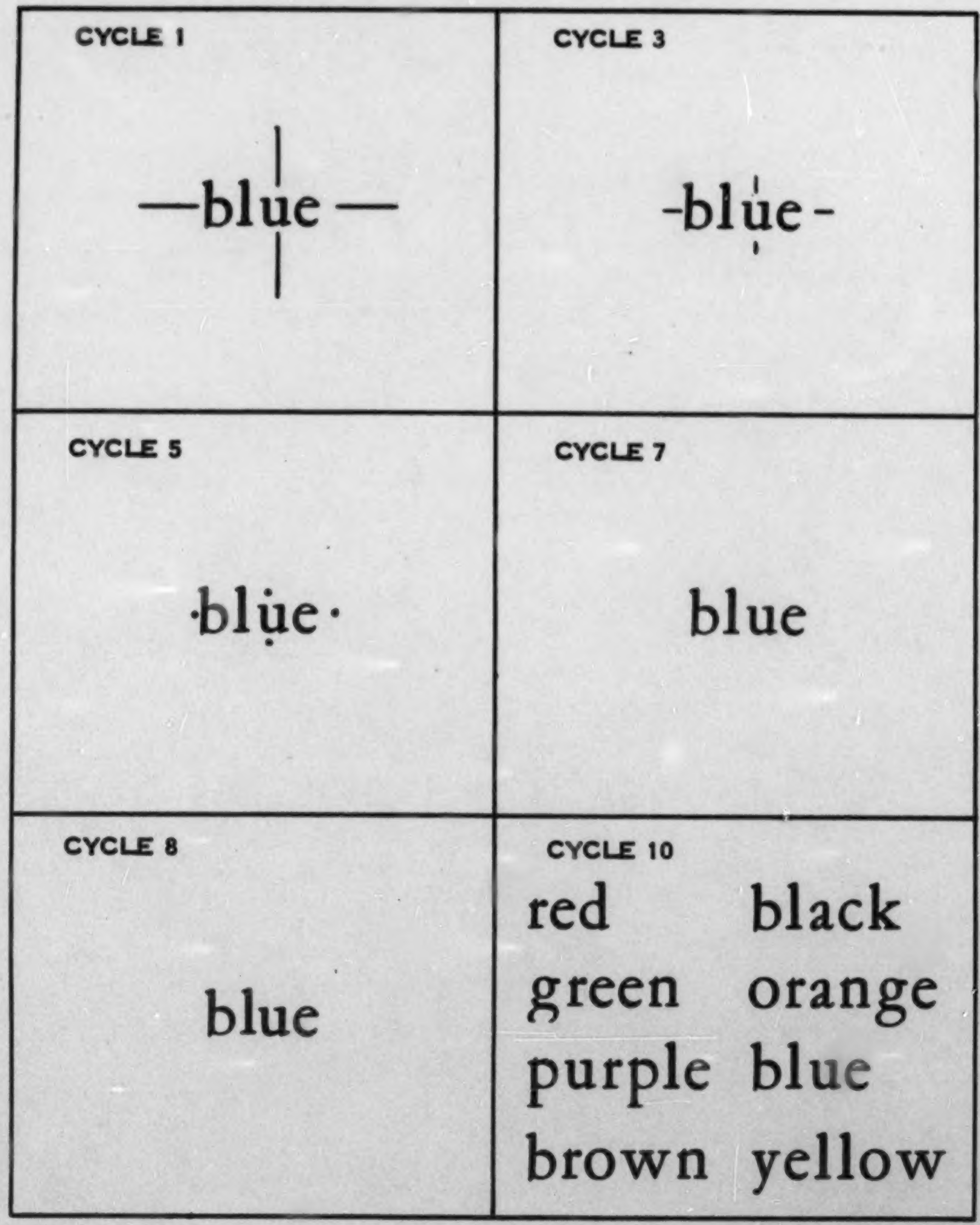
---

[2] The version of the program reported here was developed as a research tool and this report is intended primarily to describe the design, development, and early tryout of the program rather than to report on any large-scale classroom application.

By pairing the actual color with the color name, and by gradually vanishing the color prompt, it was hypothesized that the verbal response would transfer from the color to the name of the color. To fade a color prompt would lead to difficulties in the mechanics of printing the program and so vanishing was chosen as the most feasible technique. It remains a question, however, whether fading or vanishing is the superior method. Figure 9.1 shows how the color prompt for the word blue was vanished throughout the programmed sequence. In each cycle of the program, each of the eight words appeared once. In the early cycles of the program only one word appeared on each page and four appropriately colored lines radiated out from the printing of the word. All words were printed in black except as noted below. In each successive cycle the prompting color lines decreased in size until they appeared only as dots and finally were omitted altogether. A manuscript style of printing was used which seemed to be a good compromise between the various printing systems taught in the first grade. The prompting of each of the cycles is described in Figure 9.2. Each frame in the program was on a separate sheet of 5 x 8" paper. Each frame consisted of only one color word with or without the appropriate prompting lines. Each word appeared only once in a cycle containing all eight words and there were 11 cycles in the entire program. Thus, in going through the program a child reads each of the eight words a total of 11 times. The order in which the words appeared was varied from cycle to cycle. The entire program consisted of a total of 74 frames.

Figure 1

Frames from the Learning Color Names program showing prompt vanishing for the word "blue" at different points in the program. (Note: frames reduced in size by about one-third.)



CYCLE 1

—blue—

CYCLE 3

-blue-

CYCLE 5

·blue·

CYCLE 7

blue

CYCLE 8

blue

CYCLE 10

red
green
purple
brown

black
orange
blue
yellow

Table 9.2

The Method of Prompt Fading Used in the "Reading Color Names" Program

Cycle 1:	Colored radiating lines 1 inch long.
Cycle 2:	Colored radiating lines 1/2 inch long.
Cycle 3:	Colored radiating lines 1/4 inch long.
Cycle 4:	Colored radiating lines 1/8 inch long.
Cycle 5:	Four colored dots, replacing the lines at the points where the lines originated.
Cycle 6:	Two colored dots, placed at the beginning and end of the word.
Cycle 7:	No lines or dots. Each word printed in its appropriate color.
Cycle 8:	Two letters of each word printed in the appropriate color, the rest of the letter printed in black.
Cycle 9:	All words in black, no prompts.
Cycle 10:	Cycle ten consisted of one page with all of the eight words printed together on the page in black.
Cycle 11:	Cycle eleven was the same as cycle ten except that the words were printed in a different order. Cycles ten and eleven were included as pre-tests, and post-tests of reading ability.

As previously implied, the form of the program described in Table 9.2 was considered a transitional experimental form. Several variations of the program led up to the format reported here. In addition, several variations have been made as a result of experience with it. For further work the omission of cycle eight is now suggested because this cycle was found to occasionally lead to confusion. It also seems desirable to break the program up into two parts so that learning could be done in two sessions rather than in one. With two parts, four words could be learned completely in the first session and the remaining four words in the second session. Although learning all eight words with the Color Names program commonly took only a few minutes, the children were intensely active and seemed to tire quickly. All versions of the program had a cover with the title "Learning Color Names." The letters of this title were printed in the eight colors of the program and were used as a pre-test to test the ability of the child to identify the colors.

Program Development and Tryout

The program was administered by an adult who sat with the child and generally turned the pages attempting to adjust the rate of frame presentation to the child's readiness to respond. The first frame of the program was shown to the child and he was asked simply to "read this word" or "read what this word says." Further directions were usually necessary, however, to draw attention to the colored lines, but after the first frame no further urging seemed to be needed. From the second frame on, the child went through the program calling out the name of the color represented by the word and the associated color prompts. Liberal use was made of such reinforcers as "right" and "good" and repeating the color word after the child.

The first version of the program used all typewriter capitals in printing the letters and used a single vertical colored line drawn through the center of the words. None of the letters in any of the words ever appeared in anything but black. This first version was tried out on two six-year-old children. These children could read all eight of the words at the end of the program. Both children demonstrated great "personal involvement" at the completion of the program and asked to be allowed to have the book to practice with. (The only copy of the first trial version now occupies an honored place in a child's toybox.) The program was given to these children approximately six weeks after they had started first grade, where they had had some exposure to color words.

Two versions of the program were then prepared. These versions were similar to the version described in the preceding table, except that one version used no colored letters as in cycles seven and eight, i.e., the frames in cycles seven and eight were simply unprompted "criterion" frames printed in black letters. Initial tryout of these two versions was carried out with first grade children selected by their teacher from a reading readiness group. These children were considered to be slow learners and their I.Q.'s ranged between 75 and 90. In this pilot study only three children were tested on each of the two versions. The children's reading knowledge of the color words before taking the program was almost nonexistent despite previous special training on the words in the reading readiness class. A few of the children could pick out the words yellow and red. As before, an adult sat with the children, turning pages and supplying reinforcement for correct answers. Error rates were extremely low, but latencies were often long. Running time for a single child was between four and six minutes for the total program of 75 frames. After finishing the program, three of the six children could read all eight of the words and two could read six words without prompting. One child was excused early in the program because of lack of attention and general negativity. Both versions seemed to work equally well on this small sample.

This version of the program reported in Figure 9.1 was also tried out in November of the school year with teacher-selected kindergarten children. The teacher was asked for above average children. Three boys and three girls went through the program at this time. Two of the boys learned all eight words at the end of the session, and one boy and the two other girls were unable to read any of the terminal behavior words. There appeared to be a good relationship between the child's ability to identify the colors readily, a behavior which most children learn in kindergarten, and the success of the program. The two girls who failed to learn any of the words had considerable trouble in naming the colors on sight at the beginning of the program. None of the kindergarten children could read any of the words at the beginning of the program.

Approximately 15 children have been run at various times since the kindergarten tryout. These subsequent tryouts have used students with differing amounts of experience with words and letters. These children have been in the 5 1/2 to 6 1/2 year-old age range. The results obtained lead to the conclusion that the program functions as an adequate teaching device when, and only when, the control stimulus to be transferred or extended from (a color) elicits the naming response to be attached to the new stimulus. This is the initial required repertoire for which the program was constructed. The existence of the initial repertoire constitutes a condition of "readiness" for the program. A further point to be made is that in the course of the usual kindergarten and first grade schooling the age range at which the student is ready for the color names program varies a great deal and the period during which the program has optimal utility seems to be rather short for a given child. The child should be able to identify all eight colors promptly with very little or no confusion between colors. He should also have a reasonably favorable history of interpersonal contact. It is worth further note that the program works best when individually administered by an adult who supplies prompt and consistent reinforcement or feedback. The program, however, might be usable with a group of students as "seat work." It seemed desirable, also, for the student to have a copy of the program for some period of time after taking it if he showed any interest in practicing on his own. Another version of the program which we have found extremely useful is one which splits the eight color words into groups of four as described earlier.

Discussion

The color names reading program as it has been described is an example of how learning to sight-read may be programmed efficiently according to certain basic notions of learning theory. This should not be taken to mean that we recommend sight reading as a method of teaching a large reading vocabulary. Indeed, other kinds of training such as phonics seem to have much more to recommend them, since such training provides a student with behaviors that facilitate learning new words on his own. As a small first experience in reading, however, some sight reading training provides the student with an additional experience with words that takes a minimum of learning time and which often gives the student "the feeling" of being able to read and thus encourages his reading progress. Many teachers, in fact, make a point of teaching color names early in first grade because they are easy to teach and because they are useful as directions in seat work with workbooks that require coloring.

The discriminative transfer technique using literal or behaviorally equivalent prompts has wide generality to a variety of subject matters. Such a programming procedure could be used wherever it is desirable to attach or transfer an existing set of discriminations to a new set of stimuli. The authors have thought of a number of subject matters in which such a situation is met and perhaps the reader will think of others. Using the literal prompt system, a number names reading program for numbers up to 100 has been constructed. Children in first grade usually learn first to count, then to verbalize the numbers in response to the sight of the arabic numerals, and finally to read the printed number words. Such sequential acquisitions serve as the basis for sequential transfer programs. An early frame from a number reading might look like this:

7
7 seven 7
7

Much of elementary reading training could be carried on with the simple paradigm of the literal prompt discriminative transfer program. Words such as "dog," "cat," and "house" all evoke vocal responses when the real objects, or pictorial representations of real objects, are present. These representations offer vanishable prompts for reading programs. Verbs, too, can often be prompted in a variety of ways; for example, a series of human stick figures engaged in various activities such as sitting, running, eating, etc. could be used as prompts in association with the written verb. Through instructions the child taking such a verb program could be given a "set" to responding with verbs. Such stick figure prompting has, in fact, been employed in language texts (Richards, Metcalf, and Gibson, 1950). Discrimination transfer programming offers a start or describes one aspect of reading programming but, it must be pointed out, by no means offers the complete answer.

Another area, similar to reading education in children, is adult foreign language training as implied above. It would be interesting, for example, to prepare a color names program in, say, Russian for an English speaking adult. The adult would, of course, have to learn first to say Russian words in response to the colors themselves either through tutoring, self-instruction or explicit programming.

Shorthand is among the subject matters which might incorporate this type of programming. (The reader is reminded that most popular shorthand systems are copyrighted.) In this subject the student's existing responses to word sounds must be transferred (or extended) to a novel set of symbols. The same is true for the learning of code symbols and abbreviations. The technique has also been used with a beginning electricity program designed to teach recognition of the symbols used in circuit diagrams. In this program, for use with adults, the responses which already exist to words such as "resistor" and "speaker" are transferred to the representing diagrammatic symbols. In general, the discriminative transfer procedure permits the extension of responses from non-verbal to verbal stimuli or from verbal to non-verbal stimuli depending upon which has been established previously.

The color names program could easily be put into machine format. In a simple mechanical machine the student could respond to each frame by putting a check above, or a circle around, one of eight colored dots; i.e., an eight choice multiple-choice response. After the student responded to a frame the correct answer, revealed by a lever-push, could be shown as a dot of the correct color. Automating the program would probably not insure that a correct spoken response become attached to each word but there is little reason to think that it would fail to occur if the required initial repertoire were present and if the material were carefully programmed. More elaborate machines, of course, could be constructed that would permit the use of other response modes, such as a button push, and other techniques of fading and vanishing. Such a degree of automation would have to be made for reasons other than those for which the present program was designed. Research in programming was, in this case, adequately served by putting the program into booklet form.

The use of a human, as in the color names program, as a teaching machine or as part of a teaching system has certain advantages. When the program is used in booklet form the teacher functions as a page turner and reinforcement mediator. A short demonstration program such as this could be used by parents and beginning teachers as a training device which provides a "feel" for programming and for the use of reinforcement. In this respect it is probably better than demonstrations or films of animal learning experiments.

The teaching of reading by the use of such obvious prompts as color and form may seem like a very mechanical process. Even if a child should learn to read the words taught in the "Learning Color Names" program, the teacher might be well justified in asking if the child "really understands" the words he responds to. In one sense the answer to such a question is *no,* because use of the program shows, for example, that although the program teaches word recognition it does not teach the child how to spell. In fact, spelling a word is a behavior which is very different from that of reading a word. It is also true that a child who has used this program may not be able to read color names when he sees them printed in different scripts, nor will he "understand" that green is a product of the mixture of blue and yellow. However, if such understanding is desired it should be behaviorally defined and brought into the programmed sequence.

References

Angell, D. & Lumsdaine, A. A. *Prompted and unprompted trials versus prompted trails alone in paired associate learning.* Pittsburgh: American Institute for Research, October 1960. OSR Scientific Report No. 2 AIR No. 31460-IR- 129.

Richards, I. A., Metcalf, Ruth C., & Gibson, Christine. *Spanish through pictures.* New York: Pocket Books, Inc., 1950.

Skinner, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.

Skinner, B. F. The programming of verbal knowledge. In E. Galanter (Ed.), *Automatic teaching*: The state of the art. New York: Wiley, 1959. Pp. 63-68.

PRECEDING PAGE BLANK-NOT FILMED

Chapter 10

Symbols for Use in the Editing of Programmed Learning Sequences[1]

At the University of Pittsburgh Programmed Learning Laboratory an extensive effort was undertaken to prepare programmed learning sequences as experimental vehicles, as teaching aids, and as didactic exercises for teacher-programmers. The professional staff of the Programming Laboratory included three psychologists who, in addition to their assignments in writing frames, carried the responsibility of editing the frames produced by as many as 15 non-psychologists. These non-psychologists were for the most part, quite unfamiliar with behavioral psychology at the beginning of their training as programmers. Most of the programmers working at the Laboratory were public school teachers, college teachers, or graduate students in English, mathematics, etc. The behavior of such programmers requires considerable "shaping" through selective reinforcement of their verbal behavior, the most important form of which consists of the frames they write. A great deal of this selective reinforcement and discrimination training must be given by the editor who is responsible for criticizing the programming efforts of a beginning programmer. This training is often confounded by the fact that many of those who come to the Laboratory for training come with at least some indifference to behavioral psychology in general.

Frame editing, then, from the psychologist's point of view, is not simple criticism of style, usage, and grammar. The editing task required that many frames be gone over in a short period of time (so that feedback to the programmer will not be overly delayed) and that the editor have available a standard system of notations for use at points where differential reinforcement seems indicated. The purpose of this paper is to report, and to call for comment and improvement on, a system of editorial notation. The chief usefulness of the rather elaborate system of editorial notes has been in didactic programming, i.e., programming which a teacher does in order to improve his teaching. In the production of programs for their own sake the editor will find use for only a small number of the symbols presented in this paper.

---

[1] The original paper was prepared by Julian I. Taber, Halmuth H. Schaefer, and Robert Glaser and was issued as a separate project report by the University of Pittsburgh, December 1960.

One of the best ways to train programmers would be to have a program on programming. Some group will no doubt write such a program in the near future basing it upon the rapidly increasing understanding of programming.[2] There seems now to be enough "teachable" material on the subject to make such an effort of some lasting value (Lumsdaine & Glaser, 1960; Glaser, 1960; etc.). Early in a programmer's training there is little else to do, at the present time, except spend considerable time with him. Readings in programming may serve as a basis for discussion between trainee and psychologist, but the detailed discussion period, on an individual trainee basis, remains unavoidable. Even in the discussion period there is often little guarantee that whatever shaping takes place will generalize to frame writing behavior. It seems advisable to have him start writing frames early in his training. Frame writing time may increase rapidly early in training if there is a means at hand to supply differential reinforcement for this type of behavior. After an initial orientation and after production of frames has started, it is possible to keep in touch with the behavior of the programmer through occasional personal contacts and through the increasingly important mode of editorial comment.

Each student programmer is provided with a sheet of editorial symbols with their meanings listed. An effort is made to insure his correct interpretation of each symbol which requires or assumes that the programmer has some familiarity with certain behavioral principles. The use of such symbols permits the editor a more rapid review of frames and saves considerable time which would otherwise be spent in writing out long comments. The first draft of most frames is written on slips of paper which permit and encourage revision, addition, and deletion of frames without the effort of complete rewriting. Although some more venturesome writers prefered to write the first draft with several frames on one page, or even to write the first draft of a course unit on a single Foringer scroll, a writer is less likely to change the material so written. With a set of standard and precise editorial symbols it is possible to insert an editorial comment or direction at exactly the spot on an individual frame where it is needed. It should be pointed out that the symbols are used primarily as training aids which place continued emphasis upon the behavioral principles which underlie programming. The symbols are used as positive reinforcers to point up and strengthen good programming practices.

---

[2] Since the time this report was written, a short program on programming by Susan Meyer Markle has appeared. Its title is, "A Programed Primer on Programing," and may be obtained from The Center for Programed Instruction, New York. A similar program, "The Principles of Programed Learning," is also available from the Grolier Society, New York.

There are a number of ways of imparting at least a beginning understanding of behavioral principles to the programmer trainee. One way is the traditional assignment of a book such as Keller and Schoenfeld (1950) or Keller (1954) followed by discussion. Reading assignments, however, may be dispensed with, at the cost of greater teaching time, in favor of a lecture-discussion approach. Another good method is to have the student programmer complete a programmed course in basic psychology such as the Holland-Skinner program. Some may question the value of any exposure to the basic notions of behavioral psychology and an occasional programmer does seem to have some "knack" for programming that permits him to turn out good work without explicit training. Training in the vocabulary of the editor, however, facilitates communication and seems to provide greater flexibility of technique. Such flexibility is highly desirable, since the emphasis in the Laboratory was the application of learning principles to program construction, and the discovery of new programming techniques which these principles may generate.

The principles which the programmer is first required to command are relatively few. An effort is first made to have the trainee substitute the words "stimulus" and "response" for the words "question" and "answer." When these words are established, it is possible to transfer and broaden their meanings to include more of the connotations of the behavioral psychologist. It is an easy step, then, to explain and illustrate the various means by which stimuli may be brought to control behavior; i.e., successive approximation, generalization, discrimination training, etc. Discrimination training of single responses and the notation of the discriminative repertoire seem to be terms which provide the programmer with useful tools. For example, in specifying the goals of any specific educational attempt it is of little help in programming to ask the programmer to teach an "understanding" or "knowledge" of the material. Often it seems that traditional approaches to teaching treat something like "understanding the water cycle" as a single response unit when, upon analysis, it can be shown to consist of a discriminative repertoire containing many separate stimulus-response units which are more feasible teaching blocks with which to work.

When a programmer analyzes the subject matter to be taught with the intention of listing as many instances of discriminative responses as possible, he is automatically suggesting to himself the kinds of frames he must write. If the programmer sees that the verbal response "evaporation" must frequently be emitted in response to the word "heat," he at once knows that a number of frames must elicit the response in the presence of the stimulus. In turn, the word "heat" is a response that must come under the control of other words and contexts and so still other frames are suggested. Since most programming is of verbal material and since most verbal material is emitted in the form of response chains or sequences, a knowledge of response chaining is extremely valuable. This understanding of chaining requires a shift of thinking from that of traditional educational practice. For example, in most school-room situations, where there is a list of material to be learned, the teacher starts at the beginning of the list and works forward to the end of the list; children learn the alphabet from "a" to "z" or to count from one to ten. Laboratory studies on chaining, however, start with the last or terminal behavior and work back to the initial response. There is much to be said concerning this difference in the approach to response chaining, and the implications for programming and for appropriate experimentation are great. To summarize, the programmer should have an understanding of the following terms before the editorial symbols we have found useful may be used on his work: (a) stimulus; (b) response and response repertoire; (c) control stimulus, stimulus control, and discrimination training; (d) conditioning; and (e) generalization.

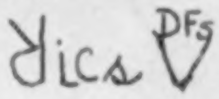
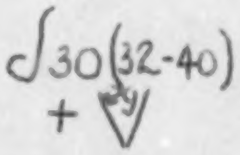
Editorial Symbols

Я	The wrong response is likely to be emitted in response to this frame.
$\not{R}$	The stimulus is too general and it is possible that no response will be emitted to this frame.
-R→	A response previously learned to this stimulus material may intrude and produce a wrong response; existing responses are likely to interfere with the desired response here. (This symbol often finds a use if the programmer fails to realize that a student brings many "association values" to a program with him.)
-R-	The range of possible correct responses is too wide.
$\overset{\text{eg}}{\vee}$	Add an example or example frame(s) before you try to evoke this response.
$\overset{\text{ru}}{\vee}$	Add a rule or rule frame (statement of the generalization being taught.)
$\not{Q}$	There is no need to state this frame as a question. Let the response emerge in the context rather than as an answer.
Q	Insert a query or test for terminal behavior. (Equivalent to saying, "Fade all prompts.")
₰	Delete
!	Good use of existing response repertoire.
C	Consider the use of chaining here; or C! = good use of chaining.
(m...n)	Rewrite frames from m to n.

?	Subject matter seems questionable. Have an expert check.
>	Split into separate frames as marked.
W	Too wordy.
RR	Too many responses. Split up into several frames or limit to one response.
|R|	This calls for a trivial response. Response is already at high strength. Prompting is too strong.
D	Discrimination, discrimination training. (Seldom used by itself. For example, "Df's" used with a caret means insert a series of discrimination frames or "discrimination series." A discrimination series would present, randomly, stimuli which should and should not evoke a particular response in such a way that the error rate remains low but a terminal discrimination is arrived at.)
G	Generalization. ("Gf's" used with a caret means to insert a series of generalization frames. A "generalization series" might, for example, be added to a program to insure that the student has explored the full range of total implications of some subject-matter principle.)
Wm	Warm-up or review frames. (A programmer may assume that more is retained from a previous section of a program than seems warranted or the programmer may fail to seed in review frames.)
+	Add
s	Stimulus
r	Response

cs	Control stimulus, a stimulus which now evokes a given response; an effective stimulus.
$\overline{cs}$	Stimulus which is to become a control stimulus; a potential control stimulus.
vg	Very good
g	Good
a	Adequate
i	Inadequate
fig	Figure
ul	Underline
cap	Capitalize
□	Panel
E	English, faulty use of language.

All of the symbols listed above may be used by themselves to represent words or phrases. In combination, they seem to eliminate the need for most of the handwritten comments necessary for frame editing. A little practice with the use of these symbols will illustrate their versatility. Below are given some symbol combinations and the verbal expressions they represent.

Symbol	Editorial Comment
	"The student will probably give the wrong response here because of an inadequate control stimulus. I suggest you insert a series of frames to establish the discrimination first."
	"Delete frame 30 and rewrite frames 32 to 40 using (more, better, different) examples."

$\not{R}\!\rightarrow\!\not{R}\ \bigvee^{wmfs}$ "You may get interference from past learning to produce a wrong response. I suggest you add a series of warm-up frames."

$W \overset{3}{>} RR$ "Too wordy. Split this frame up into at least three frames. You are calling for too many responses in one frame."

The symbols presented have been found useful in rendering editorial comments quickly and precisely. They have been developed and used in writing comments of frames in general science, bookkeeping, consumer education, elementary arithmetic, logic, set theory, psychology, meterology, and other subject matters. As new topics are programmed, new symbols may be needed. With use, some of the symbols may be found to be dispensible. It has been found valuable to train programmers to use behavioral language and the symbols seem to help stress this point. In any event, it is hoped that the list will be improved as work progresses and that other laboratories will offer their comments, suggestions, and experiences. Others contemplating the use of editorial symbols are urged to consult a list of standard proofreader's symbols, contained in many good dictionaries, since some of these may occassionally be of use. The following appendix show some examples of the use of the symbols with specific frames in a program on psychology.

APPENDIX

Some Examples of the Use of Editorial Symbols in Criticizing Rough Draft Frames in an Eighth Grade Psychology Program

---

S-1. An organism ue is any animal. A mouse is both an ______ and an animal.

Sue

R-1. .organism

S-2. Another name for animals which have organs ng such as the heart, liver, and brain is ______.

2>

R-2. organism

S-3. [① Some small animals do not have any organs such as the stomach and brain,] [but they have other parts which do the same jobs. We would still call such small animals ______ ② .]

2>

R-3. organisms

S-4. "Organisms" is ~~are~~ another name for ______.

E

R-4. animals

S-5. (Most organisms are able to move about or at least do a certain number of things. Psychologists are people who are interested in the beh ______ of organisms. -R-

R-5. behavior

S-6. Your ______ is what interests psychologists the most.)

-R- or R

R-6. behavior

S-7. You, your pet dog (if you have one), mice, fleas (if you have any), cats, elephants, cows, fish, crows, pigs, donkeys, sheep, alligators, and rhinoceroses are all ______.

R-7. organisms

S-8. ______ is another name for the things that organisms do.

i.s

R-8. behavior

References

Glaser, R. Principles and problems in the preparation of programmed learning sequences. A report prepared under cooperative Research Program Project Number 691 (9417); The United States Office of Education, Department of Health, Education, and Welfare.

Keller, F. S. Learning: reinforcement theory. New York: Random House, 1954.

Keller, F. S., & Schoenfeld, W. N. Principles of psychology. New York: Appleton-Century-Crofts, 1950.

Lumsdaine, A. A., & Glaser, R. (Eds.) Teaching machines and programmed learning. Washington, D. C.: National Education Association, 1960.